

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:28:20 ; Search time 56 Seconds
(without alignments)
2176.142 Million cell updates/sec

Title: US-10-706-424-10
Perfect score: 7901
Sequence: 1 MNSQDSITELSLPKGGGA.....WFTVNEDENTAAEVKKVM 1474

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5 COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6042	76.5	1481	2	US-09-251-645-14
2	6031.5	76.3	1476	2	US-08-817-514A-4
3	4581.5	58.0	1485	2	US-08-851-567B-32
4	160.5	2.0	1426	2	US-09-492-709A-340
5	155.5	2.0	658	2	US-09-252-991A-24910
6	155.5	2.0	2315	2	US-09-543-681A-5434
7	154	1.9	4630	2	US-09-091-609-2
8	154	1.9	5215	2	US-09-105-537-2
9	153.5	1.9	1028	2	US-09-543-681A-7181
10	144	1.8	1377	2	US-09-711-164-467
11	143	1.8	798	2	US-09-489-039A-10045
12	142.5	1.8	1183	1	US-08-447-031A-2
13	140	1.8	979	1	US-08-346-455B-38
14	140	1.8	979	2	US-08-977-221-38
15	140	1.8	979	2	US-09-483-831B-70
16	140	1.8	979	4	PCT-US95-06613-38
17	139.5	1.8	646	2	US-09-902-540-10353
18	139.5	1.8	1577	1	US-08-793-824-2
19	139	1.8	2200	2	US-09-796-575-2
20	138	1.7	1529	2	US-09-215-694-1
21	138	1.7	1529	2	US-10-109-310-1
22	137	1.7	2628	1	US-08-570-311-14
23	136.5	1.7	990	1	US-08-232-540-2
24	136.5	1.7	990	1	US-08-428-949A-2
25	136.5	1.7	990	1	US-08-428-948A-2
26	136.5	1.7	990	1	US-08-428-946-2
27	136.5	1.7	990	4	PCT-US95-04656-2

28	136.5	1.7	1013	1	US-08-233-008A-8	Sequence 8, Appli
29	136	1.7	1481	2	US-10-050-763-1	Sequence 1, Appli
30	135.5	1.7	667	2	US-09-328-352-4294	Sequence 4294, Ap
31	135.5	1.7	2199	4	PCT-US95-11684-2	Sequence 2, Appli
32	135.5	1.7	2199	4	PCT-US95-11684-2	Sequence 2, Appli
33	135	1.7	788	1	US-08-346-455B-36	Sequence 36, Appli
34	135	1.7	788	2	US-08-977-221-36	Sequence 36, Appli
35	135	1.7	788	2	US-09-483-831B-36	Sequence 36, Appli
36	135	1.7	788	4	PCT-US95-06613-36	Sequence 36, Appli
37	134.5	1.7	1277	2	US-09-397-885-3	Sequence 3, Appli
38	134.5	1.7	1277	2	US-09-969-362-3	Sequence 3, Appli
39	133.5	1.7	1626	2	US-09-252-991A-23805	Sequence 23805, A
40	133	1.7	1665	2	US-09-543-681A-4476	Sequence 4476, Ap
41	131.5	1.7	1244	2	US-09-543-681A-6274	Sequence 6274, Ap
42	129.5	1.6	1092	2	US-09-275-608-3	Sequence 3, Appli
43	129.5	1.6	1092	2	US-09-423-126-5	Sequence 5, Appli
44	129.5	1.6	1190	2	US-09-252-991A-21474	Sequence 21474, A
45	127.5	1.6	1302	2	US-09-902-540-14853	Sequence 14853, A

ALIGNMENTS

RESULT 1

US-09-251-645-14

; Sequence 14, Application US/09251645

; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.

; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: C0C1963/A

; CURRENT APPLICATION NUMBER: US/09/251,645

; CURRENT FILING DATE: 1999-02-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 14

; LENGTH: 1481

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

; US-09-251-645-14

Query Match 76.5%; Score 6042; DB 2; Length 1481;
Best Local Similarity 76.1%; Pred. No. 0;
Matches 1127; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy	1	MNSQDSITELSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGVAPAFITLNNY 60	Sequence 8, Appli
Db	1	MNSQTSFMTLSLPGKGGALTGMGEALTPTGPDGMAALSPLPISAGRGVAPAFITLNNY 60	Sequence 1, Appli
Qy	61	SGAGNSPFLGWCDCNVMTIRRTTFGPHYDETDITFLGPEGEVLVVA-----DQPRDES 114	Sequence 2, Appli
Db	61	SGTGNSPFLGWCDCNVMTIRRTSTGVPNYDETDITFLGPEGEVLVVAALAGQADIRSES 120	Sequence 36, Appli
Qy	115	TLOGINIGATFTVTGYSRSLSHESFRLRYEQPTTKTGTDFTWLIYSPDQVHLLGKSPQAR 174	Sequence 36, Appli
Db	121	SLQGINLGMTFTVTGYSRSLSHESFRLRYEQPTTKTGTDFTWLIYSPDQVHLLGKSPQAR 180	Sequence 36, Appli
Qy	175	ISNPSQTTQTAOWLLEASVSRGBOIYYOVRAEDDTGCEADEITHLQATQRYLHVYV 234	Sequence 2, Appli
Db	181	ISNPLVNNQTAQWLLEASVSRGBOIYYOVRAEDDTGCEADEITHLQATQRYLHVYV 240	Sequence 1, Appli
Qy	235	GNRTASETLPGLDGSAQDWLFLYVDFYGERSNLKTTPAFSTTGSWLCRCQDRFSRYE 294	Sequence 2, Appli
Db	241	GNLTASEVFTPLNGDDPLKSGWLFCLVDFYGERKNSLSEMPFPKATSNWLCRCQDRFSRYE 300	Sequence 2, Appli
Qy	295	YGFEIRTRLCRQVLYMYHHLQALDSKITEHNGPTLVSRLLNLYDESAIASTLVFVRVGH 354	Sequence 2, Appli

Qy 595 L P G S Q P A T E N P A Q V Y L A D L D G S G P T D L I Y V H T N R L D I F L N K S G N G P A E P V T L R F P P G L 654
Db 601 L P G S Q S A S F N P R V H L A D L D G S P A D L I Y V H A D R L D I F N E S G N G P A K P F T L S F P D G L 660
Qy 655 R F D H T C O L O M A D V O G L G V A S I L S V P H M S P H H W R C D L T N M K P W L L N E M N N M G V H H T L Y 714
Db 661 R F D T C Q L O A D V O G L G V V S I L S V P H W A P P H H W R C D L T N A K P W L L S E T N N N M G A N H T L Y 720
Qy 715 R S S Q F W L D E K A A A L T T G Q T P V C Y L P P I P H T L M O T E D E I S G N K L V T L T R Y A R G A M D G R 774
Db 721 R S S V Q F W L D E K A A A L T G T P V C Y L P P V P V H T L M O T E D E I S G N K L V T L T R Y A H G A M D G R 780
Qy 775 E R E R G F G Y V E Q T S H Q L A O G N A P E R T P P A L T K W Y A T G L P I D N A L S T E W R - D D Q A P A 833
Db 781 E R E R G F G Y V E Q T S H Q L A O G N A P E R T P P A L T K W Y A T G L P A V D N A L S A G Y W R G D K Q A P A 840
Qy 834 G F S P R F T W O N K D V L T P E D N S R Y W F N R A L K G O L L R S E L Y G L D D S T N K H V P V T V E F R 893
Db 841 G F T P R F T L W K G K D V L T P E D H N L Y W L N R A L K G O P L R S E L Y G L D G S A Q Q I P T V T E S R 900
Qy 894 S Q V R L Q H T D S R Y P V L M S V V E S R N Y H Y E R I A S D P Q C S Q N I T L S S D R F G Q P L K O L S V Q Y P 953
Db 901 P Q V R Q L O D G A T V S P V L M A S V V E S R Y H Y E R I I S D P Q C N Q D I T L S S D L F G Q P L K Q V S V Q Y P 960
Qy 954 R R O Q P A I N L Y P D T L P D K L L A N S Y D D Q Q R L B L T Y Q Q S W H H L T N N T V R V L G L P D S T R S D I 1013
Db 961 R R N E P T T N P Y P D T L P T F A S S Y D D Q Q L L R L T C R Q S W H H L I G N E L R V L G L P D G T R S D A 1020
Qy 1014 F T Y G A E N V P A G A N L E L L S D K N S L I A D D K P R E Y L C O Q K T A Y T D G O N T P L O T P T R Q A L I A 1073
Db 1021 F T Y D A K Q V P D G L M L E T I C A E N S L I A D D K P R E Y L N Q R T F Y T D G N Q T P L K T P T R Q A L I A 1080
Qy 1074 F T E T V F N Q S T L S A F N G S I P S D K L S T L T L E Q A G Y Q T N Y L P P R T G E D K W V A H H G Y D Y G T 1133
Db 1081 F T E T A V L T E S L S A F D G G I T P D E L P G I L T Q A G Y Q E P V L F P R T G E N K W V A R Q Y D Y G T 1140
Qy 1134 A A Q F W R P O K Q N T O L T G H I L I W A N Y C V V Q T R A D A G L T S A K Y D W R F L P V Q L T D I N D 1193
Db 1141 E A Q F W R P V A Q R N S L L T G R W L K W D T H Y C V I T Q T Q D A A G L T V S A N Y D W R F L T P T Q L T D I N D 1200
Qy 1194 N O H L I T D A L G R P I T L R P W G T E N G W T G Y S S P E K A S P S P D V N A A T E L K P L P V A Q C V 1253
Db 1201 N V H L I T D A L G R P V T Q R F W G L E S G A T G Y S S E K P S P P N D I D T A I N L T G L P V A Q C L V 1260
Qy 1254 Y A P E S W M P V L S Q K T F N R L A E D W K L Y N A R I I T E D G R I C T L A Y R W V Q S Q K A I P Q L I S L L 1313
Db 1261 Y A P D S W M P L F S Q E I F N T L T Q E Q E T L R D S R I I T E D W R I C A L T R R W L Q S Q K I S T P L V K L L 1320
Qy 1314 N G P R L P H S L T L T T D R Y D H D P E Q I R Q O V V F S D G F G R L L Q A A A R H E A G M A R Q E N D G S L 1373
Db 1321 T N S I G L P P H N L T L T T D R Y D R S E Q I R Q O V A F S D G F G R L L Q A S V R H E A G E A W Q R N D G S L 1380
Qy 1374 I I N V Q H T E N W A V T G R T E Y D N K G Q P I R T Y Q P Y E L N D W R Y V S D S A R Q E K E A Y A D T H Y D P 1433
Db 1381 V T K V E N T K R W A V T G R T E Y D N K G Q I R T Y Q P Y E L N D W R Y V S D S A R - - K E A Y A D T H Y D P 1438
Qy 1434 I G R E I K V I T A G W F R R T L F T P W F T V N E D E N T A A E 1468
Db 1439 I G R E I R V I T A G W L R Q S Q Y P F W F T V S E D E N T A A D 1473

RESULT 3

US-08-851-567B-32
; Sequence 32, Application US/08851567B
; Patent No. 6528484

GENERAL INFORMATION:

; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: firench-Constant, Richard

; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapiinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosiences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-851-567B-32

Query Match 58.0%; Score 4581.5; DB 2; Length 1485;

Best Local Similarity 57.6%; Pred. No. 0;
Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

Qy 1 M Q N S Q D S I T E L S L P K G G A I T G M E A L T P T G P D G M A L S L P L P I S A G R G V A P A F T L N Y N 60
Db 1 M Q D S P E V S I T T L S L P K G G A I N G M E A L N A A G P D M A S L S L P L S T G R G T A P G L S I Y S 60
Qy 61 S G A G N S P F G L G W C N V M T I R R T H F G V P H Y D E T D T F L G P E G E V L V A - - - - - D Q P R D E 113
Db 61 N S A G N G P F G I G W C G V M S I S R R T Q H G I P Q Y G N D D T F L S P Q C E V M N I A L D Q G Q P D I R Q D V 120
Qy 114 S T L Q I N I G A T F T V T G Y R S L E S H F S R L E Y W Q P K T - - T G K T D F W L I Y S P Q G V H L G K S P 171
Db 121 K T L Q G V T L P I S Y T V T Y Q A R Q I L D F S K I E Y W Q P A S G Q E G R A - F W L I S T P D G H L I L G K T A 179

```
QY 172 QARISNPQTOTAOMLLEASVSRGEOIYQYRAEDDTGCEADELTHLQATAQRYLHI 231
Db 180 QACLANPONDQQAQMLLEETVTPAGEHVSQYRAEDEAHCDNEKTAHNVTQAQRYLVQ 239
QY 232 VYVGNRTASBTLPGDGSAPSQADWLFYLVFDYGERNNLKTTPAFST-TGSWLCRQDRF 290
Db 240 VNYGNIKPQASLFLVDNAPAPPEMLFHLVFDHGERDTSLHTVPTWDAQTAQMSVRPDI 299
QY 291 SRYEYGFETRRLCRQVLMYHHLQALDSKITHRNGPTLVSRLLINVDSEAIASLTIVFR 350
Db 300 SRYEYGFETRRLCRQVLMYHHLQALDSKITHRNGPTLVSRLLINVDSEAIASLTIVFR 359
QY 351 RVCHGEDGNVVTLPPELEAYQDPSRHHAWQPMVDLANENAIQWOLVDLKGEGPLGL 410
Db 360 QLSHESDGRVPTQPLELAWQRFDEKIPWQRFADLNFNSQRYQLVDRGEGPLGL 419
QY 411 YQDKGAWYRSARQALGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQQLDWITGPG 470
Db 420 YQDKGAWYRSARQALGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQQLDWITGPG 479
QY 471 LRGYHSORPGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530
Db 480 IRGYHSORPGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539
QY 531 GFAGKDVVQSGDITLVPVPCADPRKLVAFSDVLGSGOAHLEVSATKVTCPNLRGRFG 590
Db 540 GWRKGDVPOSTGTLVPTGTDARKLVAFSDVLGSGOAHLEVSATKVTCPNLRGRFG 599
QY 591 QPITLPGFOPATFPAQVYLVADLQSGPTDLYVHTNRDLIFLNKSGNGFAEPTVTRF 650
Db 600 QPITLPGFOPATFPAQVYLVADLQSGPTDLYVHTNRDLIFLNKSGNGFAEPTVTRF 659
QY 651 PEGRLFDHTCOLQADVQGLVGLVSLILSVPHMSPHHWRCDLTNNKPLWLNEMNNMGVHH 710
Db 660 PEGRLFDHTCOLQADVQGLVGLVSLILSVPHMSPHHWRCDLTNNKPLWLNEMNNMGVHH 719
QY 711 TLAYRSSQFWLDEKAALTTGQTPVCYLPFPFHLTLMTQTEDEISGNKLVTLRYARGA 770
Db 720 TLAYRSSQFWLDEKAALTTGQTPVCYLPFPFHLTLMTQTEDEISGNKLVTLRYARGA 779
QY 771 WDGRERFRGFGVYQVQDSDHQLAQGNAPERTPPALTKNYATGLPVIDNALSTEYWR-DD 829
Db 780 WDGRERFRGFGVYQVQDSDHQLAQGNAPERTPPALTKNYATGLPVIDNALSTEYWR-DD 839
QY 830 QAFAGSPRFTWQ--DNKOVPLTPEDDNRYPFNRAKCOLLASELYGLDDSTNKVPY 887
Db 840 QAFAGSPRFTWQ--DNKOVPLTPEDDNRYPFNRAKCOLLASELYGLDDSTNKVPY 896
QY 888 TVTEFRSQRRLQHTDSRYPVLMSSVVSRYHYERIASDPQSONITLSSDRFGPQLKQ 947
Db 899 TVTEFRSQRRLQHTDSRYPVLMSSVVSRYHYERIASDPQSONITLSSDRFGPQLKQ 956
QY 948 LSVQYPRQOPAINLYPDTLPDKLANSYDDQORQLRLTYQSSWHHLTNNTVRVLGLPD 1007
Db 959 VDTAMPREXPAVNPYPTLPETLFDSSYDDQQLLALVQKNSWHHLTNGENWRGLPN 1018
QY 1008 STSRDIFTYGAENVPAAGLNLELLSDKNLSIADDPREYLGQKQATYDQNTPLQPT 1067
Db 1019 AQRDVVYDRSKIPTREGISLEILLKDLGLADEKAAYVQLGQQTFTYTAGQAEVLEKPT 1078
QY 1068 RQALIAETEIVFNQSTLAPNGSIPSDKLSLTLBOAGYQCTNYLFPRTGEDKVMVAHG 1127
Db 1079 LQALVAFQETAMMDTSLQAYEGVIEEQELNTALTQAGYQOQVAFNFRSPVMAARQ 1138
QY 1128 YTDYGTAAQFWRPQKQNTQLTGKITLIWDAANYCVVQVTDAAGLTTSKAYDWRFLFPVQ 1187
Db 1139 YTDYGTAAQFWRPQKQNTQLTGKITLIWDAANYCVVQVTDAAGLTTSKAYDWRFLFPVQ 1198
QY 1188 LTDINDNQHLITLDALGRPITLRFWGTENGQHTQYSSPEKASFPSPDVNAIELKPLP 1247
Db 1199 LTDINDNQHLITLDALGRPITLRFWGTENGQHTQYSSPEKASFPSPDVNAIELKPLP 1255
```

```
QY 1248 VAQCVYAPESWMPVLSQKTFNRL---AEQDWKLYNARIITEDGRICCTLAYRRWVQSOK 1304
Db 1256 VAQCVYAVDSWMPVLSLSQSQEAEALWAQRAAHMITEDGKVCALSGKRGTSQON 1315
QY 1305 AIPOLISLNNGRPLPPHSLTLTTRDYHDHPQOQIRQOVVPSDGFGRLLQAAARHAGMA 1364
Db 1316 LATQLISLASIPRLPPHVLGITTRDYSDPQOQHQQOQTVSFDGFGRLQLQSGARHESGDA 1375
QY 1365 RORNEGSLIINVQ-----HTENRWAVTGRTEYDNKGOPIRTYQPYFLNDRYVNSDA 1418
Db 1376 WORKEGGLVVDANGVLSVAPDTRWAVSGRTEYDDGQPVRTYQPYFLNDRYVNSDA 1435
QY 1419 RQEKAYADTHVYDPIGREIKVITAKGWFRRTLFTPMFTVNEDENTAA 1467
Db 1436 RD--DLFADTHLYDPLGREYKIVITAKKYLREKLYTPWFVSEDENTAS 1482
```

RESULT 4

```
US-09-492-709A-340
; Sequence 340, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: ZyeKind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
US-09-492-709A-340
```

Query Match 2.0%; Score 160.5; DB 2; Length 1426;
Best Local Similarity 19.0%; Pred. No. 0.00045;
Matches 255; Conservative 152; Mismatches 446; Indels 487; Gaps 65;

```
QY 406 LPGLLYQDKGAWYRSARQALGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQQLDW 465
Db 70 LPFILSRYSYRTKTPAPVGVFPG--WK--APSDIRLQLRDDGLILNDNGRSIHPE 124
QY 466 ITQPG-----LRGHSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLS 513
Db 125 PLLPGEAVYSRSSEMLVRGKAAQPDGH-----TLARLWALPPD 165
QY 514 LVLIQPKSVRLYANTRDGEFAK-----KDVQSGDITLVPVGDAPRKLAVAFSDVLG-- 564
Db 166 IRL-----SPHLYLATNS--AQGPWILGWSERVPGAEVDLPALPPYRVLTGMDRFGRT 219
QY 565 -----SQAHLEVVSATKVTCPNLRGRFGQPIIT 594
Db 220 LTYRREAAGDLAGEITGVTGDAGREFRLVLTQQAAREBARTS-----SLSSDSSRPLS 274
QY 595 LPQF--SQPATEPN-----AOVYLADLQSGPTDI-----IYVHTNRDLIFLNK 638
Db 275 ASAFPDTLPCTBTPGDRGIRLSAVLMLH-DPAYPESLPAPLVRYTYTEAGELLAVYDRS 333
QY 639 GNGFAEPTVTRFPPEGLRFDHTCQLQADVQGLVGLVSLILSVPHMSPHHWRCDLTNNKPL 698
Db 334 NTQV-----RAFTYD-----AQHFGWVAHYAGRPEN---RYDYDTGR---- 370
QY 699 LNEMNMMGVHHTLYRYSRSSQFWLDEKAALTTGQTPVCYLPFPFHLTLMTQTEDEISGN 758
```

371 VVEQLNAGLSYRYLYEQ-----DRITVTDLSNREV-----LHT-----EGGAGL 411
QY 759 KLVTTLRYAGANDGRERFRFGYVQTSQHLAQGNAPERTPPALTCKNWTATGLPVID 818
Db 412 KRUVKELADGS-----VTRSGYDAAGRLTAQTAAGRTTEY--GLNVWS 454
QY 819 NALS-----TEYWRDD--QAFAGFSRFTTWQDNKQVPLTPEDDNRWYFNRAKLG 867
Db 455 GDITDIITPPGRETGFYNDGNQUTAVSPDGLSREYDEPGRLVSETSR-----SG 507
QY 868 QLLRSELGLDSDTNKHPVYTVTFBRSQVRRLQHTDSKYPLVMSVVSRESNYHYERTASD 927
Db 508 ETVA--YRYDDA-HSELPAITTDATGSTRM-----TWS-----RYQLLAF 546
QY 928 PQCSQNT-LSSDFGQPLKQLSVQYPRQOPAINLPDPLDKLLANSYDDQORQLRLT 986
Db 547 TDCSGYQTRYEYDRFGQ-----MTAVHREBEGISLYRRYDNRGLTSVKDAQRETRYE 599
QY 987 Y-----QQSSWHHLTNTTV-----EVLGLPDSTRS-DI 1013
Db 600 YNAAGDLTAVITPDCNRSETQYDAGKAVSTTQGLTRSMYDAGRAVISLTNENGSHSV 659
QY 1014 FTYGAEN--VPAGLN-----LELLSDKNSLIA-----1039
Db 660 PSYDALDRLVQCGGFGDGTQRYHYDLAGKLTQSEDEGLVILWYDESDRITHRTVNGEPA 719
QY 1040 -----DDKPR-----EVLGQOKT--AY 1054
Db 720 EQMQYDGHGWLTDIHLSEGRVAVHYGYDDKGLTGECQTVENPETGELLWQHTKHAY 779
QY 1055 TD---GQNTTLPLOTPRQ-----ALIAETTVFQSTLS--AFNGSI 1092
Db 780 NEOGLANRVTPDSLPFVWELTYGSGYLAKMGLGTPLVEYTRDLRHRETVRSGMAGSN 839
QY 1093 PSOKLSTTLBOAGYQOQNTNYPRTGEDKVKVAHHDYDGTAAQFWRPQ--KQNTQLTG 1150
Db 840 AAYELTSTYYPAGLOQSHL-----NSLVYDRDYGWSDNGDLVRISGPRQTRYGYGATG 894
QY 1151 KI-----TLWDANYCVVVQTRDAAG-----LTTSAKYDMRFLTP 1185
Db 895 RLESVRTLAPDLOIRIPVAT-DPAGNRLPOPELHPDSTLTWPDNRIAEADAHVYRDEY 953
QY 1186 VOLT DIN-----DNQHLITL-----DALGRPIT 1208
Db 954 GRLTEKTRIPAGVIRTDERTHHYDSQRLVYTRYIQHGEPLVESRYLYDPLGRMA 1013
QY 1209 LRFWGTENGKMTGYSS---PEKASFSPSPDVNAALIELKKPLPVAOCQ-VYAPESWMPVL 1263
Db 1014 KRWRRER-DLTGHWLSRKPFEVTWYWGDRLLTVQ-----TDTRIQTVEPGSFTPLI 1068
QY 1264 SOKTFNRLAEQDMQKLYNARIITBGRICTLAYRRWVQSOKAIPQLISLNNNGRPLP-PH 1322
Db 1069 RVETENGEREKA-QRRSLAETLQEGS-----ENGHVVFPA 1104
QY 1323 SLTLTTRYDHPDQOQTRQOVVDFSGDGRLLQAAARHAGMARQNE--GSLIINVQHT 1380
Db 1105 ELVRLDLRL-----EEIRADRVSSESRAWLAQCGGLTVEQ-LARQVPEYTPARKAHLHYC 1159
QY 1381 ENR-----WAVTGRTEYD-----NKGOPIRTYQPYFELNDWRYVNSDSARQ 1420
Db 1160 DHRGLPLALISEDGNATWS-----AEDWGNQLNEENPHVYQYRPLPGQOH-----DE 1209
QY 1421 EKEAYADTH-VYDPI-GREI 1438
Db 1210 ESGLYNHRHYDPLQGYI 1229

RESULT 5

US-09-252-991A-24910
; Sequence 24910, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24910
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24910

Query Match 2.0%; Score 155.5; DB 2; Length 658;
Best Local Similarity 20.1%; Pred. No. 0.00031;
Matches 113; Conservative 61; Mismatches 155; Indels 233; Gaps 31;
QY 356 QDGNVVTLPPLFLAYQDFSPRHHAHQPMVDVLANFNAIQWQLVDLKGGL-PGLLYQ-- 412
Db 123 REGDNVTLVRNRNLAEDTS---IHWHGIIILPANMDGVP---GLSFEGIAPGGLYEYR 173
QY 413 ----DKGAWMYRGAORL-----GEIGSDA-----VTHEKQPLSVIP 445
Db 174 FKYRONGTYWYHSHGSLQEQAGVYGALVIDAREPEPFSYDRDYVVLSDMSDEKPORILA 233
QY 446 SLOSNAS-----LVDINGDQOLDWVIT-----GF-----GLRGY-- 474
Db 234 KLKQSDYNNFHKRTVGDFFIDVSANG---WAATLADRKMAEMKMSPTDLADVSGTYT 290
QY 475 ---HSQRPGCSWT-----RFTPLNAL-----PVEYT 497
Db 291 YLLNGQPPDGNWTGLPRGPKRLRFVNASMSYFVRI-PGLKMTVVAADQHQVPEVSD 350
QY 498 HPAQALADLMAGSLDLVLIGPKSVRLY-----ANTRDGAQKDVQVQSGDITLIPVQAD 552
Db 351 ELRIAVAETVD-----VIVEPGRRAYTLFAOSMDRSGYARGTLALAEGLSAPVPTPD 403
QY 553 PRKLVAFSDVLGSGQAH-----LVEVSATKVTCPNLG-----585
Db 404 PRPLIGMDDMGCMGDHGAHGAAATPASEMDHSGKMSGMDMGMDHSGKMGMDMGMDH 463
QY 586 -----RGRFGQFITLPGFSQPATE-FNPAQVYLADLDGSGPTDLIYVHTNRLDIFLNK 637
Db 464 SKMAGMDHSGMGNG-AMPMQSHPEASEDGNP---LVDMQTWTPTPKL-----ADPGLGL 512
QY 638 SGNG-----PAPVTVLRF--PGLRFDHTCQLQOMADVQGLGVASLILSVPHMSPHHWRCD 690
Db 513 RDNRRRLVTYAD-LRSRFADPDGREGPRTIELHTG-----HMEKFANFSFD 557
QY 691 ---LTNWKPLWANE-----MNNMNGVHTLRYRSSSQFMDLDEKAAALTGTQTPVCYL 739
Db 558 GIKFSOAEPLRLTYGERLRLVNDTMMTH-----587
QY 740 PFPHTLMQTEDEISGNKLV 761
Db 588 PIHLGMM-SLDEDE-QGNFLV 607

RESULT 6

US-09-543-681A-5434
; Sequence 5434, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706


```
QY 661 QLMADVQGLGVASLILSVPHMSFHHW--RCDLTNMK-----PWLLENMNNMVGHHLLRY 714
DB 2442 VLVGTGGGLGA-----LLARHLVERHCVRELLLSRGADAPGAADLGEDLAGAEVAF 2497
QY 715 RSSQFWLDBKAAALT--GQTVCVLPFPIHLMOTE--TEDEISGNKLVTTLR-YARG 769
DB 2498 AAADAADRESLARAIATVPAHEP---LTAVVHTAGVVDDATVEALTPERLDAVLRPKVDA 2554
QY 770 AMQGRE--REFRGFGVYEQDTSQHL-----AQGN--APERTPPALTKWYATGLPVIDNAL 821
DB 2555 AWWLHETKDLRLDADFVSSVSGIVGTAGQAYAAANTGLDAAHRAATGLAA--TSL 2612
QY 822 STEYWRDQAFAG--SPRETTWQDNKDVLTPTEDDNRYSWFNRALKQGLLRSELYGLD 879
DB 2613 AWGLWDGTHMGCTLGAADLARSRAGITPLP-----LQGLALFPAAVARD 2660
QY 880 STNKHVPYVTFRQVRRLOH--TDSRYPLMSSVVESRNYHYERIASDQCSQNTLS 937
DB 2661 AL--LVP-----AGLRPTAHRGTGQPPALWRGLVRAPRAARTAAE-----A 2702
QY 938 SDRFGQPLKOLSVOYP--RRQOPAINLYPDLTPDKLLANSYDDQORQLRLTYQSSWHHLT 996
DB 2703 ADTTGGWLSGLAAQSPERRSTAVTLVTGVAD----- 2735
QY 997 NNTVRVLGLPDSRSDIFTYGAENVPAGGLNLELLSKNSLIADDKPREVLGQOKTAYTD 1056
DB 2736 -----VLGHADSA-----AVGAER-----SPKD-----LGFDSLAGE 2763
QY 1057 GQN-----TTPLOTPTRQALIAFTTTFVFNOSTLSAFN-----GSPDSKLT--TLEQAGY 1106
DB 2764 LRNLNAATGLRLPA-----TTFVDHPSPAALASHLLAQVPGLEKGTAAATVVAE 2814
QY 1107 QQTNYLFPRTGEDKWVAHGYDYGTAQF-----WRPQOSNTQLTG-KITLI 1155
DB 2815 RGASFQDRATDDDFIAIV-----GMACRYPGGVSSPEDLWRLVAEGTDAISEFPVNRG 2867
QY 1156 WD-----ANYCVVQTRDAAGLTSAKYDMRE--LTPVQLTDINDNOHL--- 1197
DB 2868 WDLSELYDPDPESKGYTYC-----REGGLEAGDFDAAPFGISPREALVMDPQORLLE 2922
QY 1198 ITLDALGR-----PITLFWGTENGKMTGYSSPEKAS 1299
DB 2923 VSWEALERAGIDPSSLA--GSRGGVYVGAAGHSYAS 2956

RESULT 9
US-09-543-681A-7181
; Sequence 7181, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7181
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7181

Query Match 1.9%; Score 153.5; DB 2; Length 1028;
Best Local Similarity 18.0%; Pred. No. 0.001;
Matches 204; Conservative 124; Mismatches 356; Indels 449; Gaps 52;

QY 19 GAITGMEALPTGPDGMAAL-----SLPLPISAGRGVAPATLNNYAGNSPGLG 71
DB 162 GSATGKGLDITAPDQAGTLFFDQIIMSVPD---NRWPIPDYQIPYVNNVNTMYSKN 218
```

```
QY 72 WDCNVMTIRRRTHFGVPHYDETDITFLGPEGEVLVADQPRDESTLQGINLGATFTVTGYR 131
DB 219 WSALLM-----YDQMFTAHP---TLNFATEFRDD-----Q 246
QY 132 SRLSHFSRLBYWQPKTKTKDFWLIYSPDQVHLHGKSPQARISNPSOTTOTAQMLLEA 191
DB 247 PEVASIYQRFYYQ-----GVSRRKKITAEMDKMLALM----- 280
QY 192 SVSSRGEOIYYQRAEDDTGCEADEITHHLQATAQRYLHIVVYGNRTASETLPLGLDGAP 251
DB 281 -----KKLALQSHA----- 294
QY 252 SQADMFLYLVFOYGERSNMLKTPAPSTTGSWMLCRODRFSRYEYGEPIRRLCROVLAY 311
DB 295 AXA-----LDHPNQNFIKVEGVPE-----ETQK----- 319
QY 312 HHLQALDLSKITEHNGTLV-----SRLLNVDESAIASTLV-----FVRVGH 354
DB 320 ---ALLDANMLRDVGKTLQTAIYLRSRSLSAIDRKKLETLVLLGTRYVLEQGFTRGSGY 376
QY 355 EQDGNVVTLPPLLEYAD-----PSPRHHAHQPMQMDVLANFNATQRMQLVDLKGELP 407
DB 377 Q-----IIT-----HVGQYTRLEFDAMFIGRH-----ILAKHNLAPTO----- 410
QY 408 GLLYQDKGAWYRSARQLGEIGSDAVTWBKMOPLSVIPSLQSNASLVLDINGD---QQLDW 464
DB 411 -----QAMWYNATGRI-----FEKDNEIVDANVDILNTQLQW 443
QY 465 VITGPGLRGYHGORPDG---SWTRFTPLNALPV-----EYTHPPRAQLADLM 507
DB 444 MIKSLMLPDYQORQOQALAQLOSLWLNKKTILSSKGAGGFKSDGSIPIHSHQYPAYAKDAF 503
QY 508 GAGLSDLVILGPKSVRL---YANTPDGPAKGVQVQSGDITLPVPG-----ADP 553
DB 504 GGLAPSVYALSHSPFLSSPAHARLKVLLKRIYTKETQIPLVLSGRHPTGLHKISIDP 563
QY 554 RKLVPASDVLGSOQHLVFSATKVTCPNLGR-----GRFGQBITLPGFS 599
DB 564 FKWMALAGTPDGKQ---ELDTTLAAYAKLANKDSFEGIKAEENEPVGAWMYASMAIQ 619
QY 600 QPATEFNPAAQVYLADLDGSGPTDLI-----YVHTNRLDIF---LNKSG----- 639
DB 620 RRASITAPQOQSWLAIAARGFS--RYLVGNESYENNNRYGRYLQYQGLRILIPADLTSGFSHA 678
QY 640 ---NGFAEPVTLRFPPEGLRPHDTCOLQMAVQGLGVASLILSVPHMSPHHWCDDLTKMK 695
DB 679 GNDWNPYGTTHILP---YDE--LEAKUSQPSAGIEEMLLSTQRYSG----- 722
QY 696 PMLLENMNNN---MGVHHTLYR-----SSSQFWLDBKAAALTG---OTPV 737
DB 723 ---ANTLNNMFPAMKLGHSKYQOQSLRANKSYFLFDRNVIALGSGIENNDKQHTTETT 779
QY 738 YLPFPPIHTLMQTEDEISGNKLVTTLRYARGAWDGRERFRGFGVYEQDTSQHLQAGNA 797
DB 780 LFQFAVPEKL-----QSIILNGKKVNL-----GTQLTLNNA 810
QY 798 PERTPALTKWYATGLPVIDNALSTEYWRDD-----QAFA-----CFSPRFTTWQD 844
DB 811 DTLIDPAGNLYKLAAG--QTVFESYQKQYSDVDRNSOOTQLFATAVISHGKAPK----- 863
QY 845 NKDVPLTPEDDNRYSWFNRALKQGLLRSELYGLDDSTNKHVPYVTFEFSRQVRLQHTDS 904
DB 864 -----NANYEYATAIEAQ-----DNKAPETV-----LQHNQ 891
QY 905 RYPLMSSVVESRNYHYERIASDQCSQNTILSSDRFGQPLKOLSVOYPRQ-----QP 958
DB 892 LHAVK-DKITQEBGYAFFN-ATEVNSSQALLSSD---SP-TWVMVKQKQQLTSLVNP 945
QY 959 AINLYPDLTPDKLLANSYDDQORQLRLTYQSSWHHLT-----NNTVRVLGL 1005
DB 946 DLNLYQGIADQI-----DNKGQVEVSYSROW---LTADPPISSTVTVKGI 991
```

```

RES001 10
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA 008A
; CURRENT APPLICATION NUMBER: US/09/711.164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-711-164-467

```

Query Match	1.8%;	Score 144;	DB 2;	Length 1377;
Best Local Similarity	18.6%;	Pred. No. 0.012;		
Matches	230;	Conservative 133;	Mismatches 403;	Indels 472; Gaps 62;

QY	406	LPGLLYQDKGAWWYSRAQRLEIGSDAVTWEKQPLSVIPSLQSNASLVDINGDQOLDWV	465
DB	70	LPFLSLRTSYGSYRKTTPAPVGSGLPG--WK--MPADIRLQLRDNTLILSDNGGRSRYFE	124
QY	466	ITGPGLRGYHSQRPDGSGWT-----RFTPL--NALPVVEY--THPRAQLAD-----	505
DB	125	HLPFGDEGY--SRSESILWVRGGVAKLDEGHRLAALWQALPEELRLSPHYLATNSQGP	182
QY	506	--LMG-----AGLSPLVLIGP--KSVRLYANTRDGFAGKQDVVQ-----SGDITLVPFGAD	555
DB	183	WLLGMCRVPEADVELPALPPYRVLTGLVDREGRQTTFHREAAGFSGEITGVTDGAW	242
QY	553	PR-KLVAFSDVLGSGQAHLEVS--ATKVTCPNLCGRFCGPIITLQFSQPAFEPNPAQV	610
DB	243	RHFLRLVLTTOAQRABEARQAISGGTSPSAPPD-----TLPGYTE-----	282
QY	611	YLADLDGSGPTDLTYVHTN--RLD--IFLNKSGNGFABPVTLRPFPEGL-----RFDHTCQ	661
DB	283	-----YGRDNGIRLSAVWLTH-----DP--EYEPENLPAAPLVRYGHWTPR	319
QY	662	LQMA-----DVQGLGVASILSVPH--MSPH-----HWRCDLTNNMKPMLLNEMNNMG	707
DB	320	GELAVVYDRSGKQVRSPTYDDKYRGRVVAHRHTGRPEIRYRSDGR-----VTBQLNPAG	375
QY	708	VHHTLRY---RSSQFWLDEKAAALITGQIPVCVLPFPIHTLWQTEFEDISGNKLVTTL	764
DB	376	LSYTYQYKEKORITITDSLDRREVLTQGE-----AGLKRYYVKK	413
QY	765	RYARGAWDGREREPGRFGYVE-QTDSHQLAQGNAPERTPPALTKNWVATGLPVIDNALST	823
DB	414	EHADGS--VTQSQPDVAGRLRAQTD---AAGRTEYSPDVVT--GLITRITPDPGRASA	465
QY	824	EYWRDDQAFAGFSRFTTWQDNKDV-----PLTPEDNSRYFWNRALKGOLLR	871
DB	466	FYNNHNNQLTS-----ATGPDGLRELRREYDELGRLIQETAPDGDITRYV-----	510
QY	872	SELYGLDDSTNKHVPYVTTFERSQVRRLQHTDSRYVJWSSVRESNHYHVERIASDPQCS	931
DB	511	-----DNPHSDLP-CATE-----DATGSRKTTWS-----RYGQLLSFTDCS	546
QY	932	QNIT-LGSDRFQ-----PLKQLSVQVPRRQOPAINLYPDTLPDKLLANSYDDOQRQ	982
DB	547	GVTTRYDHDHDFQMTAVHREBGLSQYRAYDSRGQLIAVK-----DTQCHE	591
QY	983	LRLTYQOOSWHHLTNTNTRVVLGLPDPDSTRSDIFTTYGAENVPAGGUNLELLSDKNLSLIADKK	1042
DB	592	TRVEY-----NIAGDLTAVIAPDG	610

RESIT.T 11

```

RESULT 11
US-09-489-039A-10045
; Sequence 10045, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSIS AND THERAPEUTICS
; FILE REFERENCE: 2709, 2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10045
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-10045

```

```

Query Match      1.8%; Score 143; DB 2; Length 798;
Best Local Similarity 20.9%; Pred. No. 0.0056;
Matches 137; Conservative 81; Mismatches 232; Indels 206; Gaps 34;

Qy 18 GGAITCGEALTTGPDGMAALSLPISAGRCYAPAPFTLYNNSGAGNSPFGLGWDC--- 74
Db 123 GYNIRGMESNRVGDVDGIAQ-----PNATGRGV-----GRAGLNTFGIGRDYIDP 169
Qy 75 -----NVMTIRRRTHFGV-PHYDET----- 94
Db 170 YMTGSVDIQSGATSTETANSAIGNVSRFKSADDYLRPKTSAFGYRSGYDSADRSWHN 229
Qy 95 --TFLGPEGEVL--VVADQPDESTLQGINGLATFTVTGYSRRLSEH-FGRLEYWQP--- 146
Db 230 GVTAVAGD-EFLRGLVYSRDGOETN-NSG---TVDAVPANWHSDAFSLASGIWQND 284

```


Db 731 --DKDNQDGRPEKVS-----VNLADGB---KVKILDVTS---ETNWKYEFKDLPKYD 777
Qy 1194 --TPVQLTIDINDNQHLLITLALGRPITLRFMGTEGKMTGYSPEKASFSPSDVNAAE 1241
Db 778 EGKIEYVTEHDHVXDYTDINGTITNKY-----TPGETSAIV- 816
Qy 1242 LKPLPVAQCQVYAPESMPVLSQTFNRLABQWQKLYNARIITEDGRICTLAYRRWQ 1301
Db 817 -----TKNDNNN-----QDKRPEIKVELYQ 840
Qy 1302 SOKAIPOLISLNNPRLPHSLTLTDRYDHPQRIQQQVPSDGFGRLLQAAARHEA 1361
Db 841 DKGATGK-TALINES-----NNWHTWTGLD---EKAQQQVKYTVBELTKVGYTHV- 890
Qy 1362 GMARQRNEDGLIINVOHTENRWAVTGRTEYDNK-----GQPIRTYQ 1403
Db 891 ---DNNDMGNLIVNTKYPETTSISGEKVDNDKQDQKPEKVSNNLANGKVKTLTD 946
Qy 1404 PYFLNDWRYVNDSSARQK-----EYADTHVYDPIGREIKVITAKGWFRTLTPT 1454
Db 947 VTSETNWKYBFDKLPKYDEGKKIEYTVTEDHVKYDITDINGTITNKYTPGETSATVTKN 1006
Qy 1455 WFTVNEDENDTAAEVK 1470
Db 1007 WDDNNQDGRPEIK 1022

RESULT 13

US-08-346-455B-38
; Sequence 38, Application US/08346455B
; Patent No. 573167
; GENERAL INFORMATION:
; APPLICANT: UNITED STATES OF AMERICA; DEPT.
; APPLICANT: OF HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: MOTILITY STIMULATING
; TITLE OF INVENTION: PROTEIN USEFUL IN CANCER DIAGNOSIS AND
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & PINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,455B
; FILING DATE: 28-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06613
; FILING DATE: 24-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/249,182
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/822,043
; FILING DATE: 17-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4149PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:

LENGTH: 979
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: Unknown
MOLECULE TYPE: protein
HYPOTHETICAL: No
ORIGINAL SOURCE:
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE: Liver
CELL TYPE:
CELL LINE:
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: putative autotaxin
OTHER INFORMATION: protein sequence from human liver
US-08-346-455B-38
Query Match 1.8%; Score 140; DB 1; Length 979;
Best Local Similarity 18.9%; Pred. No. 0.015;
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;
Qy 418 WYSAORLGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDGLDWI----TGPGLRG 473
Db 261 WITAKORGE-----SWN-----ILLVCHPSRAILTI-----LQWTLDPHERPSVA 305
Qy 474 YHSQPDGSMTRFTPLNALPVEYTHPRAQLADLGA---GLSDLVLIQPKSVRLYANTRD 530
Db 306 FYSEQDPFSGHKHMPFGP---EMTNPLREMHKIVGQMDGLKQLKLRVCNV-IFVETMD 361
Qy 531 GFAK-----GKDVQSGDITLVPVGA-----DPRKLVAFSDVLGSGQAH 569
Db 362 GRCHMYRTEFLSNYLTVNVDITL--VPGTLGRIRSKFSNNAKYDKAIIA----- 409
Qy 570 LVEYSATKVTCPNLPNGRGRFQPIITLPGFSQPAFEFPAQVYLADLDGSGPTDLYVHTN 629
Db 410 -----NLTC-----KKPDQHPK-----YLRQ-----HLPRLHYANNR 438
Qy 630 R-----LDIFLNGSGNGFAEPVTLRFPEGLRFDHTCOLQADVQGLG 671
Db 439 RIEDIHLLVERWHVARKPLDVYKKPSGNAFSETT-----AFDNKVNMQTVFVGYG 491
Qy 672 ----VASLILSVPHMSPHWRCDLTNNKPMLLNNNNMMGVHHTLYRVRSSQFVLDEKAA 727
Db 492 PTPKYTKVPPFENIELYNVMDLLGLK-----APNNGTHGSLNHLRTNTFRPTMPE 545
Qy 728 ALTTGQTP-VCYLPFP-----IHTLQOTETEDISGNKLVTT 763
Db 546 EVTRPNYPGIMYLOSDFDLCCTCDDKVEPKNKLDELNKLRLTKGSTBERHILYGDPAVL 605
Qy 764 LR-----YARGAWDGRERFRGFGYVEQTDSDHOLAQNAPE-----RTPPALTK 807
Db 606 YRTRYDILYHTDPESGYSEIFLMLWTSYTVSKAEVSSVPDHLTSCVRPDRVSPSPSQ 665
Qy 808 NWYATGLPVIDNALSTYWRDDQAFAP-----SP--RFTTWQNKQVPLPTEDDNN 857
Db 666 NCLA-----YKNDKQMSYGFLLFPYPLSSSPSEAKYDAFLVTNMVMPYPAKRV 712
Qy 858 RYWFENRAL-----KGQLARSELYGLDDSTNKHVPVTVTFEFSQVRRLQHTD 903
Db 713 WNYFQRLVVKYASERNGVNISGIFDYDVGHLHDEDKIKQIV-----EG 759
Qy 904 SRYFVLMSSVSESNHYERIASDPQSONITLSSDRFGQPLKQLSVQYPRROOPAINLY 963
Db 760 SSIPV-----PTHYYSIITS-----CLDFTQADKCDGLSVSSFILPHR----- 799
Qy 964 PDLPLDKLLANSYDDQORQLRLTYQQSW---HHLTNNTRVRLGLPDPDSTRSDIFTYGAENV 1021

Db 800 PD---NEESCNSD-----ESKWEELMKHHTARVDIEHLTSLDFRKTSSY 846

Qy 1022 PAGLNLLELSDKNSL 1037

Db 847 P-----EILTLYL 856

RESULT 14

US-08-977-221-38

; Sequence 38, Application US/08977221

; Patent No. 6084069

GENERAL INFORMATION:

; APPLICANT: UNITED STATES OF AMERICA; DEPT.

; APPLICANT: OF HEALTH AND HUMAN SERVICES

; TITLE OF INVENTION: MOTILITY STIMULATING

; TITLE OF INVENTION: PROTEIN USEFUL IN CANCER DIAGNOSIS AND

; TITLE OF INVENTION: THERAPY

; NUMBER OF SEQUENCES: 69

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/977,221

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/346,455

FILING DATE: 28-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/249,182

FILING DATE: 25-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/822,043

FILING DATE: 17-JAN-1992

ATTORNEY/AGENT INFORMATION:

NAME: DOROTHY R. AULT

REGISTRATION NUMBER: 36,434

REFERENCE/DOCKET NUMBER: 2026-4149US3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 979

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: Unknown

MOLECULE TYPE: protein

HYPOTHETICAL: No

ORGANISM: Human

STRAIN:

INDIVIDUAL ISOLATE:

DEVELOPMENTAL STAGE:

HAPLOTYPE:

TISSUE TYPE: Liver

CELL TYPE:

CELL LINE:

ORGANELLE:

FEATURE:

NAME/KEY:

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION: putative autotoxin

OTHER INFORMATION: protein sequence from human liver

US-08-977-221-38

Query Match 1.8%; Score 140; DB 2; Length 979;

Best Local Similarity 18.9%; Pred. No. 0.015;

Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;

Qy 418 WYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVDINGDQLDWI-----TGQLRG 473

Db 261 WITATKORGE-----SMN-----ILLVCCPSRAEILTI-----LQMLTLPDHERPSVYA 305

Qy 474 YHSORPDGSMTRFTPLNALPVEVTHPRAOLADLMGA---GLSLVLIGPKSVRLYANTRD 530

Db 306 FYSEQDPFSGHKMPGP---EWNPLREMHKIVGQMDGLKQLKLRVCNV-IFVETMD 361

Qy 531 GFAK-----GKDVQSGDITLPVPGA-----DPRKLVAFSDVLGSGQAH 569

Db 362 GRCHMYRTBFLSNLTNVDITL-VPTLGRIRSKFSNNAKYDPKAIIA----- 409

Qy 570 LVEVSATKVTCPNLCGRFGQPITLPGFSQAPATEFNPAQVILADLDGSGPTDLIYVHTN 629

Db 410 -----NLTC-----KKPDQHFKP---YDKQ---HLPKLHYANNR 438

Qy 630 R-----LDIFLNKSGNGFAEPVTLRPFEGRLRFDTHTCOLQADVQGLG 671

Db 439 RIEDIHLVERRMHWARKPLDVYKPSGNAPSEIT-----AFDNKVNMQTVFVGYG 491

Qy 672 ----VASLILSVPHMSPHWRCDLTNMKPMLLNEMNNMNVHHTLYRSSSQFWLDEKAA 727

Db 492 PTFKYTKVPPFENIELYNVMDLGLKP-----APNNGTHGLNHLRLTNTFRPTMPE 545

Qy 728 ALTTGQTP-VCYLPFP-----IHTLMQTEDEISGNKLVT 763

Db 546 EVTRPNYPGIMYLSQDFDLGCTCDKVPEPKNKLDELNKLRLHTKSGTEERHLLYGDRAVL 605

Qy 764 LR-----YARGAWDGRERFRGFGVVEQTDHQLAGNAPE-----RTPPALTK 807

Db 606 YRTRYDILYHTDPESGYSEIFLPLMTSYTSVSKQAEVSSVPDHLTSCVRPDVRSFSQ 665

Qy 808 NWYATGLPIDNALSTYWRDDQAFAGF-----SP--RFTTWQDNKDVPPLTPEDDNS 857

Db 666 NCLA-----YKNDKQMSYGFLPPPYLSSSPEAKYDAFLVTNNVMPYAPKRV 712

Qy 858 RYWFNRAL-----KGQLLRSELYGLDSTNKHVPYVTEPESQVRLQHTD 903

Db 713 WNYFQRLVKKYASERNVVISGPIFDYDGLDHTEDKIQYV-----EG 759

Qy 904 SRYPLWSSVVERNVHYERIAADPCQSONITLSSDRFGQPLKQLSVQYPRRQOPAINLY 963

Db 760 SSIPV-----PTHYYSIITS-----CLDFTQPAKDCGDLPSVSSFILPHR----- 799

Qy 964 PDLPLPKLLANSYDDQORQLRLTYQSSW--HHLTNNVTVRVLGLPDLSTRSIDIFTYGAVN 1021

Db 800 PD---NEESCNSD-----ESKWEELMKHHTARVDIEHLTSLDFRKTSSY 846

Qy 1022 PAGLNLLELSDKNSL 1037

Db 847 P-----EILTLYL 856

RESULT 15

US-09-483-831B-70

; Sequence 70, Application US/09483831B

; Patent No. 6417338

GENERAL INFORMATION:

; APPLICANT: STRACKE, MARY

; APPLICANT: LIOTTA, LANCE

; APPLICANT: SCHIEPMANN, ELLIOTT

; APPLICANT: KRUTZCH, HENRY

; APPLICANT: MURATA, JUN

; TITLE OF INVENTION: AUTOTAXIN: MOTILITY STIMULATING PROTEIN USEFUL IN

; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

```
FILE REFERENCE: 2026-4149US4
CURRENT APPLICATION NUMBER: US/09/483,831B
CURRENT FILING DATE: 2000-01-17
PRIOR APPLICATION NUMBER: 07/822,043
PRIOR FILING DATE: 1992-01-17
PRIOR APPLICATION NUMBER: 08/249,182
PRIOR FILING DATE: 1994-05-25
PRIOR APPLICATION NUMBER: 08/346,455
PRIOR FILING DATE: 1994-11-28
PRIOR APPLICATION NUMBER: 08/977,221
PRIOR FILING DATE: 1997-11-24
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 979
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Putative autotoxin protein sequence from human
OTHER INFORMATION: liver.
NAME/KEY: VARIANT
LOCATION: (860)
OTHER INFORMATION: Xaa at positions: 860, 889, 905, 911, 927, 937,
OTHER INFORMATION: 944, 950, 954, 967, and 975 represents an unknown
OTHER INFORMATION: or other amino acid.
US-09-483-831B-70

Query Match      1.8%; Score 140; DB 2; Length 979;
Best Local Similarity 18.9%; Pred. No. 0.015;
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;

Qy 418 WYRSNQRIGEIGSDAVTEKMKQPLSVIPFSQSNASLVNDINGGQLDWVI-----TGPGLRG 473
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 261 WITATKORGE-----SWN-----ILLVCCHFSRAILTI-----LQMLTLPDHERPSVYA 305
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 474 YHSORPDGSGWTRFTPLNALPVEYTHPRAQLADLMGA---GLSDLVLIGPKSVRLYANTRD 530
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 306 FYSQPDPSGKHHPFGP---EMTNPLREMHKIVGQLMDGLKQLKLRHCNVN-IFVETMD 361
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 531 QPAK-----GKDVQSGDITLPVPGA-----DPRKLVAFSVDVLGSGQAH 569
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 362 GRCHMYRTEFLSNVLTNVDDITL-VPGTLGRIRSKFSNNAKYDPKAIIA----- 409
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 570 LVEVSATKVTCPNLRGRFGQPIITLPGFSQPAFEPNPAQVYLADLGGSGPTDLYVHTN 629
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 410 -----NLTC-----KKPDQHFKE---YLKQ---HLPKRLHYANNR 438
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 630 R-----LDIFLNKSGNGFAEPVTLRFPPSGLRDHTCQLQMAADVQGLG 671
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 439 RIEDIHLVLRVHWVARKPLDVIYKKPSGNAPFSRETT-----APDNKNSMQTVFVGYG 491
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 672 ----VASILISVPHMSPHHRCDLTNMKPMLNENNNMNVHHTLRYSRSSQFWLDEKAA 727
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 492 PTFKYKTKVPPENIELYNVMDLLGLKP-----APNNGTHGSLNHLRLTNTFRPTMPE 545
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 728 ALTTGQTP-VCYLPFP-----IHTLMQTEDEISGNKLVTT 763
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 546 EVTRPNYEGIMYLSQDFDLGCTCDKVEPKNKLDELNKLHKTGSTERHLLYGDRAVL 605
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 764 LR-----YARGAWDGRERERFGYVEQTDHQLAQNAPE-----RTPPALTK 807
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 606 YRTRVDILYHTDFESGYSEIFLMPILWTSYTSVKQAEVSSVPDHLTSCVRPDVRVSPFSQ 665
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 808 NWYATGLFVIDNALSTERYRDDQAFGE-----SP--RFTTWQDNKDVPPLTPEDDNS 857
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 666 NCLA-----YKNDKQMSYGLFPYPLSSSPPEAKYDAFLVTNNVMPYAPKRV 712
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 858 RYWFNRAL-----KGQLLRSELYGLDDSTNKHVPYVTVEFRSQVRLQHTD 903
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 713 WNYFORVLVKVYASERNGVNVIISGPIFDYDVGDLHDTEDKIKQYV-----EG 759
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 904 SRYPVLMSSVVSRYHYERIASDPQCSQNTLSDFGQPLKQLSVQYPRQOPAINLY 963
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 760 SSIPV-----PTHYYSIITS-----CLDFTQPADKCDGFLSVSSFILPHR----- 799
Qy 964 PDTLPDKLLANSYDDQORQLRLTYQSSW--HHLTNNTVRVGLPDPDSTRSDIFTYGAENV 1021
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 800 PD---NBESCNSSED-----ESKWVEELMKMHTARVRDIEHLTSLDFFRKTSSSY 846
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1022 PAGGLNLELLSDKNSL 1037
Db 847 P-----EILTLYL 856
```

Search completed: December 16, 2005, 14:42:36
Job time : 64 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:26:40 ; Search time 270 Seconds
(without alignments)
3851.662 Million cell updates/sec

Title: US-10-706-424-10

Perfect score: 7901

Sequence: 1 MQNSQDSITLSLPKGGGA.....WFTVNEDENTAAEVKKVKM 1474

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_cremi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	2	Q8GF99 PHOLU
2	7285.5	92.2	1475	2	Q7N7Y2 PHOLU
3	6065.5	76.8	1476	2	Q7N7Z0 PHOLU
4	6031.5	76.3	1476	2	Q93EP6 PHOLU
5	4627.5	58.6	1485	2	Q7N936 PHOLU
6	4595.5	58.2	1485	2	Q85153 PHOLU
7	3984.5	50.4	1433	2	Q8XP56 YERPR
8	3971	50.3	1428	2	Q9F9Z2 YENTR
9	3638	46.0	1489	2	Q6QDZ4 YERPS
10	3591	45.4	1481	2	Q693A4 YEREN
11	3569.5	45.2	1496	2	Q8ZAV4 YERPE
12	3569.5	45.2	1516	2	Q8D1P6 YERPE
13	3350.5	42.4	1401	2	Q3PRN8 XENNE
14	2277.5	28.8	1447	2	Q4ZP56 PSESY
15	2274	28.8	1446	2	Q87X46 PSESM
16	1076	13.6	2439	2	Q4HWU2 GIBZE
17	837	10.6	591	1	VRP2 SALCH
18	835	10.6	591	1	P74846 SALTY
19	832	10.5	591	1	VRP2 SALTY
20	829	10.5	591	1	VRP2 SALEN
21	826.5	10.5	593	1	VRP2 SALDU
22	826.5	10.5	593	2	Q4L1X7 YENTR
23	709.5	9.0	563	2	P72421 YENTR
24	676	8.6	283	2	Q8KSN9 XENNE
25	651.5	8.2	2031	2	Q3XEC6 BURPS
26	502	6.4	1806	2	Q62N48 BURMA
27	457.5	5.8	166	2	Q6WCC1 YENTR
28	386	4.9	2217	2	Q8TP72 METAC
29	330.5	4.2	2379	2	Q72U39 LEPIC
30	326.5	4.1	221	2	Q58FC3 SALGL
31	292	3.7	2306	2	Q8F107 LEPTOSPIRA

32	289.5	3.7	2554	2	Q72QR5 LEPIC	Q72qr5 leptospira
33	279.5	3.5	2321	2	Q8F5B9 LEPIN	Q8f5b9 leptospira
34	241.5	3.1	1826	2	Q987Z7 RHIL0	Q987z7 rhizobium 1
35	202	2.6	2370	2	Q82RE3 STRAW	Q82re3 streptomyces
36	199	2.5	2060	2	Q6Q8B0 STALU	Q6q8b0 staphylococcus
37	181.5	2.3	2364	2	Q82R58 STREAM	Q82r58 streptomyces
38	180	2.3	1976	2	Q8D4R5 VIBVU	Q8d4r5 vibrio vuln
39	177	2.2	795	2	Q7NFS7 GLOVI	Q7nfs7 gloeobacter
40	174.5	2.2	1881	2	Q8TJS7 METAC	Q8tjs7 methanobacter
41	172.5	2.2	1404	2	Q7AH10 ECO57	Q7ah10 escherichia
42	172.5	2.2	1404	2	Q8XED9 ECO57	Q8xed9 escherichia
43	171.5	2.2	2566	2	Q8TSE7 METAC	Q8tse7 methanobacter
44	170	2.2	593	2	Q8GP37 PSEAE	Q8gp37 pseudomonas
45	170	2.2	2082	2	Q9ACP4 STRCO	Q9acp4 streptomyces

ALIGNMENTS

RESULT 1
Q8GF99 PHOLU PRELIMINARY; PRT; 1474 AA.
AC Q8GF99;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE TcdB2.
GN Name=tcdB2;
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., ffrench-Constant R.H.;
RT "Genomic islands in Photobacterium.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AA017202.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integral complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; EG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
SQ SEQUENCE 1474 AA; 166338 MW; 9118D4E4914683DD CRC64;

Query Match 100.0%; Score 7901; DB 2; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MQNSQDSITLSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNN	60
Db	1	MQNSQDSITLSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNN	60
Qy	61	SGAGNSPFLGWCDCNVMTIRRRTHFGVPHYDEDTFTFLGPEGEVLVADQPDDESTLQGIN	120
Db	61	SGAGNSPFLGWCDCNVMTIRRRTHFGVPHYDEDTFTFLGPEGEVLVADQPDDESTLQGIN	120
Qy	121	LGATFTVTGYSRLESFHSRLEYWQPTTKTGTDFWLYSPDQVHLLGKSPQARISNPSQ	180
Db	121	LGATFTVTGYSRLESFHSRLEYWQPTTKTGTDFWLYSPDQVHLLGKSPQARISNPSQ	180

```
QY 181 TTQTAAWLLLEASVSSRGEQIYYQYRAEDDTCBEADETHHLQATAQRYLHIVYGNRTAS 240
DB 181 TTQTAAWLLLEASVSSRGEQIYYQYRAEDDTCBEADETHHLQATAQRYLHIVYGNRTAS 240
QY 241 ETLPGLDGSAPSQADMLFYLVDFYDYGERSNNLKTTPAFSTTGGSWLRCQDRFSRYEYGFETR 300
DB 241 ETLPGLDGSAPSQADMLFYLVDFYDYGERSNNLKTTPAFSTTGGSWLRCQDRFSRYEYGFETR 300
QY 301 TRRLCQVLMYHHLQALDLSKITEHNGPTLVSRILILNYDESAIASTLVFVRVGHQDGNV 360
DB 301 TRRLCQVLMYHHLQALDLSKITEHNGPTLVSRILILNYDESAIASTLVFVRVGHQDGNV 360
QY 361 VTLPPLLELAYQDFSPRHHAWQPMQDVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWWYR 420
DB 361 VTLPPLLELAYQDFSPRHHAWQPMQDVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWWYR 420
QY 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQQLDWITGPGLRGYHSQRPD 480
DB 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQQLDWITGPGLRGYHSQRPD 480
QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGPAGKQDVVQ 540
DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGPAGKQDVVQ 540
QY 541 SGDTITLVPVPCADPRKLVAFSDVLGSGQAHVLEVSATKVTCPNIGRGRFGQPIITLPGFSQ 600
DB 541 SGDTITLVPVPCADPRKLVAFSDVLGSGQAHVLEVSATKVTCPNIGRGRFGQPIITLPGFSQ 600
QY 601 PATEFNPAAVYLDLQSGPTDLIYVHTNRDLIFLNKSGNGGPAEPVTLRPFPEGLRFDHTC 660
DB 601 PATEFNPAAVYLDLQSGPTDLIYVHTNRDLIFLNKSGNGGPAEPVTLRPFPEGLRFDHTC 660
QY 661 QLOWADVQGLGVASLILSVPHMSPHWRCDLTNWKPLNEMNNMGVHTLRYRSSQP 720
DB 661 QLOWADVQGLGVASLILSVPHMSPHWRCDLTNWKPLNEMNNMGVHTLRYRSSQP 720
QY 721 WLDEKAAALTGTGTPVCYLPFPPTHTLMQTEDEISGNKLVTLRYARGADGGRERFRG 780
DB 721 WLDEKAAALTGTGTPVCYLPFPPTHTLMQTEDEISGNKLVTLRYARGADGGRERFRG 780
QY 781 FGVEQTDSSHQAQGNAPERTPALTKWYATGLPVIDNALSTEYMRDQAFAGFSRPT 840
DB 781 FGVEQTDSSHQAQGNAPERTPALTKWYATGLPVIDNALSTEYMRDQAFAGFSRPT 840
QY 841 TWQNDKDVPLTPEDDNRSYFNALKGQLLRSELVGLDDSTNKHVPYTVTFEFSQVRRLQ 900
DB 841 TWQNDKDVPLTPEDDNRSYFNALKGQLLRSELVGLDDSTNKHVPYTVTFEFSQVRRLQ 900
QY 901 HTDSRYPVLASSVVSERNHYERTASDPQCSQNTLSSDRFGQPLKOLSVOYPRRQOPAI 960
DB 901 HTDSRYPVLASSVVSERNHYERTASDPQCSQNTLSSDRFGQPLKOLSVOYPRRQOPAI 960
QY 961 NLXPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLTNNTVRLGLPDLSTRSDIFTYGAEN 1020
DB 961 NLXPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLTNNTVRLGLPDLSTRSDIFTYGAEN 1020
QY 1021 VPAGGLNLELLSDKNSLIADKPREYLGQOKTAYTDGQNTTPTQTRQALIAFTETTVF 1080
DB 1021 VPAGGLNLELLSDKNSLIADKPREYLGQOKTAYTDGQNTTPTQTRQALIAFTETTVF 1080
QY 1081 NQSTLSAFNGSIPSDKLSLTLEQAGYQNTYLPRTGDKVVAHGYTDYGTAAQFWRP 1140
DB 1081 NQSTLSAFNGSIPSDKLSLTLEQAGYQNTYLPRTGDKVVAHGYTDYGTAAQFWRP 1140
QY 1141 QKQNTQLTGKTLIMDANYCVVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHLITL 1200
DB 1141 QKQNTQLTGKTLIMDANYCVVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHLITL 1200
QY 1201 DALGRPTTLRPMWGTENGKMTGYSSPEKASPPSPDVNAAIELKKPLPVAQCQVAPESWM 1260
DB 1201 DALGRPTTLRPMWGTENGKMTGYSSPEKASPPSPDVNAAIELKKPLPVAQCQVAPESWM 1260
```

```
QY 1261 PVLSKQTFNRLAEQDWQKLNYNARIITEDGRICTLAYRRVWQSKAIPQLISLNNGPRLP 1320
DB 1261 PVLSKQTFNRLAEQDWQKLNYNARIITEDGRICTLAYRRVWQSKAIPQLISLNNGPRLP 1320
QY 1321 PHSLTLTTRYDHDPEQIQIRQQVVPFSDGFRLLQAAARHEAGMARQNRDGSIIINVQHT 1380
DB 1321 PHSLTLTTRYDHDPEQIQIRQQVVPFSDGFRLLQAAARHEAGMARQNRDGSIIINVQHT 1380
QY 1381 ENRMWVGTGYDNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAYADTHVYDPIGREIKV 1440
DB 1381 ENRMWVGTGYDNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAYADTHVYDPIGREIKV 1440
QY 1441 ITAKGFRFTLTPFWFTVNEDENDTAAEVKKVKM 1474
DB 1441 ITAKGFRFTLTPFWFTVNEDENDTAAEVKKVKM 1474

RESULT 2
Q7N7Y2_PHOLL
ID Q7N7Y2_PHOLL PRELIMINARY; PRT; 1475 AA.
AC Q7N7Y2_
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DE Insecticidal toxin complex protein TcdB2.
GN Name=tcdB2; OrderedLocName=plu0969;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01.
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rudiok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taouit S., Bosc S., Bouraux-Bude C., Chaudier M., Charles J.-F.,
RA Dassa E., Doros R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Signier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CAE13264.1; -; Genomic_DNA.
DR Photoliat; plu0969; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSVPBPROT.
SQ Complete proteome.
KW SEQUENCE 1475 AA; 165779 MW; 24945174CDADAD22 CRC64;

Query Match 92.2%; Score 7285.5; DB 2; Length 1475;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;

QY 1 MNSQDSFISITELSLPKGGGAIQWGEALTPAGDGMALSLPLISAGRGYAPSLALYN 60
DB 1 MNSQDSFISITELSLPKGGGAIQWGEALTPAGDGMALSLPLISAGRGYAPSLALYN 60
QY 61 SGAGNPFGLGWCNVTIRRRTHFGVPHYDETDFTLGPGEVLVADQPRDESTLOGIN 120
DB 61 SGAGNPFGLGWCNVTIRRRTHFGVPHYDETDFTLGPGEVLVADQPRDESTLOGIN 120
QY 121 LGATFTVTGYSRLESHFSRLEYWQTKTGKTFWLIYSPDGQVHLLGKSPQARISNPSQ 180
DB 121 LGATFTVTGYSRLESHFSRLEYWQTKTGKTFWLIYSPDGQVHLLGKSPQARISNPSQ 180
QY 181 TTQTAAWLLLEASVSSRGEQIYYQYRAEDDTCBEADETHHLQATAQRYLHIVYGNRTAS 240
DB 181 TTQTAAWLLLEASVSSRGEQIYYQYRAEDDTCBEADETHHLQATAQRYLHIVYGNRTAS 240
```

Db 181 TTQTQAWLLEASVSPHGSEIYYQYRAEDNLDCEITNEITLHPQATAQRVLHIVYGNRTAS 240
Qy 241 ETLPGDGSAPSQADWLFPYLPVDYGERNNLKTTPAFSTTGSWLCRODRFSRYEYGFIR 300
Db 241 ETLPGDGRAPSOADWLFPYLPVDYGERNNLKMPPAFATGNWLCRODRFSRYEYGFALR 300
Qy 301 TRRLCROVLMVHHLQALDSKITEHNGPTLVSRLLIANYDESAIASTLVPRVRGHEQDGNV 360
Db 301 TRRLCROVLMVHHLQALDNKIKENNGPTLVSRLLIANYDESAIASTLVPRVRGHEQDGT 360
Qy 361 VTLPPLELAYQDFSPRHHAHQPMQMDVLANFNAIQRWQLVDLKGELPGLLYQDQKAMWYR 420
Db 361 VTLPPLELAYQDFSPFHNTDQPMQMDVLANFNAIQRWQLVDLKGELPGLLYQDQKAMWYR 420
Qy 421 SAQRLGIGSDAVTWKQOPLSVIPSLQSNASLVINDINGDGLDQDQDQDQDQDQDQDQDQ 480
Db 421 SAQRLGIGSDAVTWKQOPLSVIPSLQSNASLVINDINGDGLDQDQDQDQDQDQDQDQDQ 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVQ 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVQ 540
Qy 541 SGDTLTPVPGADPRKLVAFSVGLSGQAHLEVSATKVTCPNLCGRGFRFGQPIITLPGFSQ 600
Db 541 SGDTLTPVPGADPRKLVAFSVGLSGQAHLEVSATKVTCPNLCGRGFRFGQPIITLPGFSQ 600
Qy 601 PATEFPAQVYLAIDLGGPTDLIVYHTRNLDIPLFKSGNGFABPVTLPPEGLRFDHTC 660
Db 601 PATEFPAQVYLAIDLGGPTDLIVYHTRNLDIPLFKSGNGFABPVTLPPEGLRFDHTC 660
Qy 661 QLOWADVOGLVASLILSVPHMSPHWRCDLTNNKPMILNMMNMMVHHHTLYRRSSQF 720
Db 661 QLOWADVOGLVASLILSVPHMSPHWRCDLTNNKPMILNMMNMMVHHHTLYRRSSQF 720
Qy 721 WLDEKAAALITGQTPVCVLPPIHTLWQTEDEBISGNKLVTLRYARGADGGRERFRG 780
Db 721 WLDEKAAALVAGQTPVCVLPPIHTLWQTEDEBISGNKLVTLRYARGADGGRERFRG 780
Qy 781 FGYYEQTDSHQAQGNAPERTPPALTKRWYATGLPVIDNALSTEWYR-DDQAQAGFGSPR 839
Db 781 FGYYEQTDSHQAQGNASERTPPAMTKWYATGLPVIDNTLSTEWYRGNDAQAGFGSPR 840
Qy 840 TTMQDNKDVLTPEDDNSRYVFNALKQGLRSELYGLDSTNKHVPYVTFEPPSQVRL 899
Db 841 TTMQDNKDVLTPEDDNSRYVFNALKQGLRSELYGLDSTNKHVPYVTFEPPSQVRL 900
Qy 900 QHTDSRYVLMSSVVSERHYERIASDPQSQNITLSSDRFGQPLKQSLVQYPRRQOQA 959
Db 901 QADNQIFVLMSSVVSERHYERIASDPQSQNITLSSDRFGQPLKQSLVQYPRRQOQA 960
Qy 960 INLYPDTLPDKLANSYDDQORQLRLTYQSSWHLTNNTVRLVGLPDSRSDIFTYGA 1019
Db 961 ISPYEDTLPDELLANSYDDQORQLRLTYQSSWHLTNNTVRLVGLPDSRSDIFTYGA 1020
Qy 1020 NVPAGLNLELLSKNSLIADKREYLQOQKATYDQONTTPIQTPRQALIAFTETTV 1079
Db 1021 NVPAGLNLELLSKNSLIADKREYLQOQKATYDQONTTPIQTPRQALIAFTETTV 1080
Qy 1080 FNQSTLSAFNGSIPSDKLSLTLEAGYQOQNYLPRGTGDKVVAHGHYTDYGTAAQFWR 1139
Db 1081 FSQSTLSAFNGSIPSAQLSTLLEAGYQOQNYLPRGTGDKVVAHGHYTDYGTAAQFWR 1140
Qy 1140 PQKQNTQLTGKITLINDANYCVVQTRDAAGLTTSYAKYDWRFLTPVOLTDINDNQHILIT 1199
Db 1141 PQKQNTQLTGKITLINDANYCVVQTRDAAGLTTSYAKYDWRFLTPVOLTDINDNQHILIM 1200
Qy 1200 LDALGRPTLTPWGTENGKMTGYSPEKASPSPSDVNAAELKPLPVACQVYAPESW 1259
Db 1201 LDALGRPTLTPWGTENGKMTGYSPEKASPSPSDVNAAELKPLPVACQVYAPESW 1260
Qy 1260 MPVLSQKTFNRLAQDQWKLNARIITDGRICTLAYRRWQSOAKAIPOLISLANNPRL 1319
Db 1261 MPVLSQKALNRLAQDQWKLNARIITDGRICTLAYRRWQSOAKATPQLTSLNNGPHL 1320

Qy 1320 PPHSLTUTTRYDHDPEQIQIRQQVPSDGFGRLLQAAARHAGMARQNRNEDGSLIINVQH 1379
Db 1321 PPHSLTUTTRYDHDPEQIQIRQQVPSDGFGRLLQAAARHAGMARQNRNEDGSLIINVQH 1380
Qy 1380 TENRWAVTGRTEYDNKQPIRTYQYFLNDWRVYNSDSAROEKAYADTHVYDPIGREIK 1439
Db 1381 TENRWAVTGRTEYDNKQPIRTYQYFLNDWRVYNSDSAROEKAYADTHVYDPIGREIK 1440
Qy 1440 VITAKGWFRRTLFTFWFTVNEDENDTAAEVKCKVM 1474
Db 1441 VITAKGWFRRTLFTFWFTVNEDENDTAAEVKCKVM 1475
RESULT 3
Q7N7Z0 PHOLL PRELIMINARY; PRT; 1476 AA.
AC Q7N7Z0;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TcdB1.
GN Name=tcdB1; OrderedLocusNames=plu0961;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01.
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Tacurit S., Bours S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Berzelle S., Freysinet G., Gaudriault S.,
RA Mouline C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunat F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CAB13256.1; -; Genomic_DNA.
DR PhotocList; plu0961; -;
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integral complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR004113; Integrin alpha.
DR InterPro; IPR003284; Sal_SpV8.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
KW Complete proteome.
SQ SEQUENCE 1476 AA; 165737 MW; 06AFE0EB945B5728 CRC64;
Query Match 76.8%; Score 6065.5; DB 2; Length 1476;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;
Qy 1 MNSQDSFISITELSLPKGGGAIITGMGEALTPTGPGMAALSPLPISAGRYAPAFITLNN 60
Db 1 MNSQTSFVAELSLPKGGGAIITGMGEALTPTGPGMAALSPLPISAGRYSPSITLNN 60
Qy 61 SGAGNSPFLGWDNCNVMTIRRRTHFGVPHYDETTPLFGPGEVLVVA-----DQPRDES 114
Db 61 SGAGNSPFLGWDNCNVMTIRRRTHFGVPHYDETTPLFGPGEVLVVA-----DQPRDES 120
Qy 115 TLQGINLGAFTTVTGYSRLESFESRLEYMQPTTKTDFWLIYSPGQVHLLGKSPQAR 174
Db 121 SLQGINLGAFTTVTGYSRLESFESRLEYMQPTTKTDFWLIYSPGQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQAWLLEASVSRGEQIYYQYRAEDDTCCEADEITHLQATQRYLHIVY 234
Db 181 ISNPLNVSQTAQWLLLEASVSHGEQIYYQYRAEDTNCETDEFTAHFNATVQYLAQVHY 240
Qy 235 GNRTASTLPGDGSAPSQADWLFPYLPVDYGERNNLKTTPAFSTTGSWLCRODRFSRYE 294

Db 241 GNLTA SEVPFLNGDDPLKSGMFLCLVDFYGERKNSLSIETPPFKASSLWLCRQDRFSRYE 300
QY YGFEIETRLRCROVLAHYHLQALDSKITEHNGFTLVSRLLTNYDESIAIATSLVFRVGH 354
Db 301 YGFEIETRLRCROVLAHYHLQALDSKITEHNGFTLVSRLLTNYDESIAIATSLVFRVGH 360
QY EQDGNVVTLPPLSLAYQDPSRHHAWQPMQDVLNANFNAIQRMQLVDLKGELPGLLYQDK 414
Db 361 EDNNTVISLPLESLAYQDPSRHHAWQPMQDVLNANFNAIQRMQLVDLKGELPGLLYQDK 420
QY 415 GAWWYSARQLGBIGSDAVTWEXQQLSPVIPSQSNASLVNDINGDQGLDWVITPGGLRGY 474
Db 421 NGWWSARQAGEENNAVTGWKQLLPITPALQDNASLMDINGDQGLDWVITPGGLRGY 480
QY 475 HSQRPPGSWTRFPPLAALPVEYTHPRAQLADLNGAGLSDLVLGPKSVRLYANTRGPAK 534
Db 481 HSQHPGWSWTRFPPLDALPIEYSHPRQAQLADLNGAGLSDLVLGPKSVRLYANTRGFTQ 540
QY 535 GKDVQSGDITLPPGADPRKLVAFSDVLGSGQAHLEVEVSATKVTWPNLGRGRFGOPIT 594
Db 541 GRDVQSGDITLPPGADPRKLVAFSDVLGSGQAHLEVEVSATKVTWPNLGRGRFGOPIT 600
QY 595 LPGFSQAPATEFNPQAQVYLDLQSGPTDLIYVHTNRDLDFLNKSGNGFAPBPVTLRPPGL 654
Db 601 LPGFSQASDNFNDRVHLADLQSGPADLIYVHTNRDLDFLNKSGNGFAPBPVTLRPPGL 660
QY 655 RPDHTCOLQWADVQGLGVASLIISVPHMSPHWRCDLTNNKPKMLNEMNNMGVHTLRY 714
Db 661 RPDHTCOLQWADVQGLGVASLIISVPHMAPHWRCDLTNAKPWLLSEMNNMGVHTLRY 720
QY 715 RSSSQFWLDBKAAALTGTQPCVCLPDPPIHTLMQTEDEISGNKLVTLRYARGAWDGR 774
Db 721 RSSSQFWLDBKAAALTGTQPCVCLPDPPIHTLMQTEDEISGNKLVTLRYARGAWDGR 780
QY 775 EREFRGFGYVEQDSDHQAQGNAPERTPPALTNNVATGLPVDNALSTEYMR-DQQAFA 833
Db 781 EREFRGFGYVEQDSDHQAQGNAPERTPPALTNNVATGLPVDNALSTEYMR-DQQAFA 840
QY 834 GFSRFTTQDNKDVPITPEDNSRWYFNALQGLRLSRLYGLDSDGNKHKVYTVTTEPR 893
Db 841 GFTPRETRWKEGKDVATPENNDFNFWFNALQGLRLSRLYGLDSDGNKHKVYTVTTEPR 900
QY 894 SOVRLQHTDSRYPLVWSSVVERSNHYHRIASDPQSONITLSSDRFGQPLKLSQVYP 953
Db 901 PQVRLQHTDSRYPLVWSSVVERSNHYHRIASDPQSONITLSSDRFGQPLKLSQVYP 960
QY 954 RROQPAINLYPDLTPDKLANSYDDQORQLRLTYQSSMHLTNNTVRLGLPDRSDI 1013
Db 961 RROQPAINLYPDLTPDKLANSYDDQORQLRLTYQSSMHLTNNTVRLGLPDRSDI 1020
QY 1014 FTYGAENVAGGLENLSDKNSLIADDKPREVLGQOKTAYTQGNTPLOTTPROALIA 1073
Db 1021 FTYDAKQVPVDGLNLEALCAENSLIADDKPREVLNQQRTFTYTGKNAQLEIETROALIA 1080
QY 1074 FTETTVNQSTLAFNGSIFSDKLSLTLEQAGYQOQNTYLPRTGDKVVAHGYTDYGT 1133
Db 1081 FTETTVNQSTLAFNGSIFSDKLSLTLEQAGYQOQNTYLPRTGDKVVAHGYTDYGT 1140
QY 1134 AAQFWRPQKQNTQLTGKTLIWDANYCVVQVTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1193
Db 1141 EAQFWRPVQARNLTLLGKTLIWDANYCVVQVTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1200
QY 1194 NQHLITLDALGRFITLFWGTENGKMTGYSPKASFPSPDVNAALTELKXPLVPAQCV 1253
Db 1201 NVHLITLDALGRFVTRFWGTENGKMTGYSPKASFPSPDVNAALTELKXPLVPAQCV 1260
QY 1254 YAPESMPLVSKTFNLEAQDQWKLNYARIITEDGRICTLAYRWVQSKAPOLISLL 1313
Db 1261 YAPDSMPLVSKTFNLTQEQOTLRDLRIITEDWRIKALARRRWLQSOQASTPLVNL 1320
QY 1314 NNGPRLPHSLTLTDRYDHDPEQIRQVVFSDGFRLLQAAARHAGMARORNEGSL 1373

Db 1321 TNSIGLPPHNLTLTPDRYDRDSQQIHOQVAFSDGFRLLQASVRHEAGEMQRNODGSL 1380
QY 1374 IINVOCHENRWAVTGRTEYDNKGQIRTYQPYFLNDWRVYVNSDSAROEKEAVADTHVYDP 1433
Db 1381 VTKMETKTRWAVTGRTEYDNKGQIRTYQPYFLNDWRVYVNSDSAR--KGAVADTHVYDP 1438
QY 1434 IGREIKVITAKGWRRTLTFTPMFTVNEDENDTAAEV 1469
Db 1439 IGREIRVITAKWLRQSQYFPWFTVSEDENTAADV 1474
RESULT 4
Q93EP6 PHOLU
ID Q93EP6 PHOLU PRELIMINARY; PRT; 1476 AA.
AC Q93EP6
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE TcdB1.
GN Name=tcdB1;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrrench-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RN Trends Microbiol. 9:185-191(2001).
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., ffrrench-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RN Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAL18487.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpB.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
SQ SEQUENCE 1476 AA; 165138 MW; 8B6AC3D109911995 CRC64;
Query Match 76.3%; Score 6031.5; DB 2; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;
QY 1 MONSQFSTELSLPKGGGAIATGMEALTPGPDGMAALSLPLISAGRGYAPAFILNYN 60
Db 1 MONSQFSTELSLPKGGGAIATGMEALTPGPDGMAALSLPLISAGRGYAPAFILNYN 60
QY 61 SGAGNSPFGLGWDCNMYTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDES 114
Db 61 SGTGNSPFGLGWDCGVMARRTTSTGPNYDETDTFLGPEGEVLVVALEAGQADIRSES 120
QY 115 TLQGINLGATFTVTGVRSLRSHFSRLYWPQTKTKTDFWLIYSDGQVHLLGKSPQAR 174
Db 121 SLQGINLGATFTVTGVRSLRSHFSRLYWPQTKTKTDFWLIYSDGQVHLLGKSPQAR 180
QY 175 ISNPSQTTTQTAOMLLBASVSSRGEQIYYOYRAEDDTGCEADETHLQATAQRYLHIVY 234
Db 181 ISNPLNVNQAQWLLBASISSEHQIYYQYRAEDAGCETDELAHAPSATVQRYLQTVHY 240
QY 235 GNRATSETLPLDGSAPSQADWLFLVYFDYGERSNLKTTPAFSTTGSWLCRODRFSRYE 294
Db 241 GNLTA SEVPFLNGDDPLKSGMFLCLVDFYGERKNSLSIETPPFKASSLWLCRQDRFSRYE 300


```
QY 291 SRYEYGEIRTRCLRQVLMYHHLQALDSDKITEHNGPTLVSRILNLYDESIASTLIVFVR 350
Db 300 SRYEYGEVTRTRCQVLMYHHLQALDSDKITEHNGPTLVSRILNLYDESIASTLIVFVR 359
QY 351 RVUGHEQGNVVTLPPLLEAYQDPSRPHHAHQWPMVLANFNALQRMQLVDLKGELPGLL 410
Db 360 QLSHEPFGSPITLPLELAQWRENLKMPWQRFDAIDNFNSQORYQLVDLGRGELPGLM 419
QY 411 YQDKGAWWYRSACRLGEIGSDAVTWKMKOPLSVIPSLQSNASLVLDINGDQLDMLVITGPG 470
Db 420 YQDKGAWWYRSACRLGEIGSDAVTWKMKOPLSVIPSLQSNASLVLDINGDQLDMLVITGPG 479
QY 471 LRGYHSQRDPGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530
Db 480 IRGYHSQRDPGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539
QY 531 GFAGKDVQVSGDITLVPQADPRKLVAFSDVILGSGQAHVVEVSATKVTCPNMLGRGRFG 590
Db 540 GMRKGDVQVSGDITLVPQADPRKLVAFSDVILGSGQAHVVEVSATKVTCPNMLGRGRFG 599
QY 591 QPITLPGFSOPATEFNPQAQVYLADLQSGGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRF 650
Db 600 QPITLPGFSOPATEFNPQAQVYLADLQSGGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRF 659
QY 651 PEGRLFRDHTCOLQWADYQVGLGVALSLILSVPHMSPHWRCDLTNNKMWLLNEMNNMNVH 710
Db 660 PEGVQDNTCOLQWADYQVGLGVALSLILSVPHMSPHWRCDLTNNKMWLLNEMNNMNVH 719
QY 711 TLAYRSSQWLEKAAALTTGQTPVCYLPFPPIHTLWQTEDEISGNKULVTTLRYARGA 770
Db 720 TLAYRSSQWLEKAAALTTGQTPVCYLPFPPIHTLWQTEDEISGNKULVTTLRYARGA 779
QY 771 WDGREREFRGVVEQDTHSLQAGNAPERTPALTKNMYATGLPVDNALSTEYWR-DD 829
Db 780 WDGREREFRGVVEQDTHSLQAGNAPERTPALTKNMYATGLPVDNALSTEYWR-DD 839
QY 830 QAFAGFSPTTWO--DNKDVPITPEDDNRVYFNFRALKQGLRSELYGLDDSTNKHVPY 887
Db 840 QAFAGFSPTTWO--DNKDVPITPEDDNRVYFNFRALKQGLRSELYGLDDSTNKHVPY 898
QY 888 TVTEFSQVRLQHTDSRYPVLWSSVVSRYHYRIASDPQCSQNTLSSDRFGPGLKQ 947
Db 899 TVTESRYQVRSIPVEKETELSAWTAENRSYHYRIISDPQLSQSLRLQHDIFGQSLOS 958
QY 948 LSVQYPRQCPALNLPDITLQKLLANSYDQDQRLTYQQSSWHLLTNNTVRLVGLPD 1007
Db 959 VDIAMPREKPAVNPVPTLPEALFDSYDDQQQRLVLRQKNSWHLLTNGENRWRLGN 1018
QY 1008 STRSDIFTYGAENVVPAAGLNLLELSDKNSLIADDPKREYLGQOKTAYTDGQNTPLQPT 1067
Db 1019 AQRDVVYTRTKIPTGEGILEVLLKADGLLADEKAAYLGQOQTFYTAGQSEVTLKPT 1078
QY 1068 RQALIAFTTFFVFNQSTLAFNGSIPSDKLSLTLLEQAGYQOQNTYLPFRTEGDKVWAHNG 1127
Db 1079 LQALVAFOETAMWDDTSLQAYEGVIEQGLNTTLTQAGYQOQVRLFNTEGSESPVWVARQ 1138
QY 1128 YTDYGTAAQFWRQKQNTQITLTKITLIDANVCVVVTRDAAGLITTSYAKYDRFLTPVQ 1187
Db 1139 YTDYRDATQFWRPQTRQNSLLTGKTLTWBTHCVIITQDAAGLITTSYAKYDRFLTPVQ 1198
QY 1188 LTDINDNQHLLITDALGRPTTLRFWGTENGWMTGYSPEKASFPSPDSDVNAALTELKPLP 1247
Db 1199 LTDINDNQHLLITDALGRPTTLRFWGTENGWMTGYSPEKASFPSPDSDVNAALTELKPLP 1255
QY 1248 VAQCQVYAPESWMPVLVSKTFNRL---AEQDWKLYNARIITEDGRICTLAYRRWQSOK 1304
Db 1256 VAQCQVYAPESWMPVLVSKTFNRL---AEQDWKLYNARIITEDGRICTLAYRRWQSOK 1315
QY 1305 AIPQLISLLNNGRLPHSLTLITDRYDHPDEQIQOQVVEFSQFGLLQAAARHEAGMA 1364
Db 1316 LTIQLISLLASIPRLPHVLTITDRYDHPDEQIQOQVVEFSQFGLLQAAARHEAGMA 1375
QY 1365 RQNEEDGSLIINVQ-----HTENRWAVTGRTEYDNKGPRTYQPYFLNDWRVYVNSDA 1418
```

```
Db 1376 WORKEDGGLVNDANGALVSAATDTRWAVSGRTEYDDKGQVRYQPYFLNNRYVYVSDSA 1435
QY 1419 ROEKRAYADTHVYDPTGREIKVITAKGWFRRLTFTPWFTVNEDENDTAA 1467
Db 1436 RD-DLUFADTHYDPLGREIKVITAKKYLUREKQYTFWFIYVSEDENDTAS 1482
RESULT 6
ID O85153_PHOLU PRELIMINARY; PRT; 1485 AA.
AC O85153;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Insecticidal toxin complex protein TcaC (Toxin complex protein).
GN NameTcaC;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens.";
RL Science 280:2129-2132 (1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX Bowen D., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191 (2001).
DR EMBL; AF046867; AAC38625.1; -; Genomic DNA.
DR EMBL; AF346497; AAL18451.1; -; Genomic DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR000408; Reg chr condens.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSEVBPROT.
DR PROSITE; PS00626; RCCL 2; UNKNOWN 1.
SQ SEQUENCE 1485 AA; 16616 MW; 74228E27A9921EBD CRC64;
```

```
Query Match 58.2%; Score 4595.5; DB 2; Length 1485;
Best Local Similarity 57.8%; Pred. No. 4.1e-284;
Matches 861; Conservative 224; Mismatches 375; Indels 29; Gaps 11;
QY 1 MONSDSFITSLPDKGGGALTQMGELTPTGPDGMAALSPLPISAGRGYAPAFITLYN 60
Db 1 MQSDPEVSTLSPDKGGGALTQMGELTPTGPDGMAALSPLPISAGRGYAPAFITLYN 60
QY 61 SGAGNSPFGLGWPCNVTIRRTTHFGVHYDETDTFLGPEGEVLVVA-----DQPRDE 113
Db 61 NSAGNGPFGIGWCGVMSISRRTHGHPQYQNDTFLSPQGVMTALNDQGPDIRQDV 120
QY 114 STIQNLIGATFTVTGYRSLSHFSRLYEQWPKT--TGKTDFWLIYSPDQVHLKSKP 171
Db 121 KTLQGVTLPISTVTRTQARQILDFSKIEYMQPASQOGRGA-FWLISPPDQGLHLGKTA 179
```

```
Qy 172 QARLSNPSTOTTAQWLLLEASVSSRGEOIYYOYRAEDDTGCEADEITHHLOATAQRYLHI 231
Db 180 QACLANPONDQOIAQWLLLEBETVPAGEHVSQYRAESEAHCDDNEKTAHPNVTAQRYLVQ 239
Qy 232 VYGNRTASETLPGLDGSAPOADWLFVLPDYDYSRNNLKTTPAFST-TGSMCLCRODRF 290
Db 240 VNYGNIRPOASLFLVLDNAPPAPEBWLPHLVFDHGERDTSLHTVPTWDAGTAQMSVRPDI 299
Qy 291 SRYEGFEIRTRRLCRQVLYHHQLADLSKITEHNGPTLVSRLLINVDSEIAISTLVFVR 350
Db 300 SRYEGFEVTRRLCQVLMFHTALMAGEASTNDAPELVRLILEYDKNASVTTLITIR 359
Qy 351 RVGHEQDGNVTLPLELAYODSPRHHAHQWMDVLANFNAIORWOLVDLKGSLGCLL 410
Db 360 QLSHESDGSPTVQPLELAWOFLEKMTWQRFALDNFNSQORYQLVLRGSLPGML 419
Qy 411 YQDKGMYRQAORLGEISDAVTWKNQPLSVPSQSNASLVDINGDGLDQWVITGPG 470
Db 420 YQDRGMYKAPQOEDGDSNAVYDYKIAPLTLPNLDNASLMDINGDGLDQWVITASG 479
Qy 471 LRGHYSQRPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530
Db 480 IRGHYSQPDGKTHFTFIPALPVEYHPSPQADLTGAGLSDLVLIGPKSVRLYANQRN 539
Qy 531 GFAKGKDVQSGDITLVPQGDPRKLVAFSDVLGSGQAHLYEVSATKVTCPNLRGRFG 590
Db 540 GWRKGEDVPQSTGTLPTVTGTDARKLVAFSDMLGSGQOHLVEIKANRVTCWPNLGHGRFG 599
Qy 591 QPITLPQSPQATBNPAQVTLADLDSGPTDLYVHTNRLDIPLNKGNGFAPSPVTLRF 650
Db 600 QPLTSLGSPQENFPERLFIADISGTTDLIYAQSGSLIYLNGSGNQFDAPLTLAL 659
Qy 651 PEGRLFOHTCOLQADVOGLVASLILSVPHMSPHHWRCDITNNKPMLLNEMNNMGVHH 710
Db 660 PEGVOFQNTCOLQADVOGLVASLILSVPHHWRCDLSLTKPMLNVMNNRGAAH 719
Qy 711 TLRYSSQFWLDEKAAALTTQTPVCVLPPIHTLWCTETEDISGNKGLVTLIRYARGA 770
Db 720 TLHYSSAQFWLDEKALQTKAGKAPACVLPPIHLLMYTETQDISGNRLTSFVNYSHGV 779
Qy 771 WDGREFRGFGYVEQDTSQALQCNAPERTPPALTKNWTATGLPVIDNALSTBYWR-DD 829
Db 780 WDGREFRFGFCIKQIDTTTFSGHTAPEQAAPSLISWFAFGMDEVDQSALATYQWQADT 839
Qy 830 QAPAGFSPRTTWQ--DNKDVPLTPEDDNRYSYFNWNRALKGQLLSLYGLDDSTNKKVPY 887
Db 840 QAYSGFETRYVMDHTNQTDAQTP-NETQRMNLTRALKGQLLRTELYGLDGTDKQTVPY 898
Qy 888 TVTFERSQVRLOHTDSRYFVLWSVBSRYHYERTASDPQCSQNTILSSDRFGQPLKQ 947
Db 899 TVSESRYQVRSIPVYNKETELSAWVTAIENRSYHYERIITDPQFSQSIKLQHDIFQSLSQ 958
Qy 948 LSVQYPRQOAIINLPDTPDKLLANSYDDQORLRLTYOQSSWHHLTNNTVRVILGPD 1007
Db 959 VDIAPREKFAVNPPTLPETLFDSDYDDQQLLRLVROKNWHHLTUDGENWRLGUPN 1018
Qy 1008 STRSDIFTYGAENYPAGNLIELLSKNSLIADDPREYLCQOKTAYTDGQNTTPTLOPT 1067
Db 1019 AQORDVITYDRSKIPTGEGISLEILLKODGLLADEKAAVYLGQQOFTFYTAGAETVLEKPT 1078
Qy 1068 QOALIAFTETTVNQSTLSAFNGSIPSKLSTTLQOAGYQNTNLPFRGTGDKVWVAHHG 1127
Db 1079 LQALVAFQETAMMDTSLQAYEGVIEQELNTALTQAGYQOQVARTNRSSEPVAAARQ 1138
Qy 1128 YTDYGTAAQWRPQKQNTQTLTKITLWDANCYVVTQTRDAAGLTSYAKYDWEPLTPVQ 1187
Db 1139 YTDYGTAAQWRPQKQNTQTLTKITLWDANCYVVTQTRDAAGLTSYAKYDWEPLTPVQ 1198
Qy 1188 LTDINDNQHLITLDALGRPITLRFWGTENGQWTVYSSPEKASFPPSPDVNAAIBLKXPLP 1247
Db 1199 LTDINDNQHIVTLDALGRVITSRFGWTEAGQAAGYSN---QPFTPPDSVDKALALTGNLP 1255
Qy 1248 VAQCOVYAPESMWPVLQKTNRL---ABQDWKQLYNARIITEDGRICTLAYRNRWQSQK 1304
```

```
Db 1256 VAQCLVTVADVSWMPSSLSSQSSQSEAEALWAQLRAAHMITEDGKVCALSGRKGTSQHN 1315
Qy 1305 AIPQLISLLANGPRLPPHSLTLTTDRYDHPDPEQOIRQOVVFSFGFGRLLQAAARHEAGMA 1364
Db 1316 LTIQLISLLASIPRLPHVLIITDRYDSDPQQOHOQTVFSFGFGRLLQSSARHESGDA 1375
Qy 1365 RQNEDEGSLINQV-----HTENRVAVTORTEYDNKGQPIRTYQPYFLNDWRVYVNSDA 1418
Db 1376 WORKEDGGLVVDANGVLVSAPTDRWAVSGRTYDDKQGVRTYQPYFLNDWRVYVNSDA 1435
Qy 1419 RQKEAVADTHVVDPIGREGIKVITAKGFRRTLTFTWFTVNEDENDTAA 1467
Db 1436 RD--DLFADTHLDPLGREYKVTAKKILREKLYTFWFTVSEDENDTAS 1482

RESULT 7
Q6XP56_YERFR PRELIMINARY; PRT; 1433 AA.
ID Q6XP56_YERFR PRELIMINARY; PRT; 1433 AA.
AC Q6XP56;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE TcyF2
OS Yersinia frederiksenii.
CG Plasmid pMYL.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=29484;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=49;
RA Dodd S.J., O'Callaghan M., Ronson C.W.;
RT "Yersinia enterocolitica contains plasmid-encoded genes that show
similarity to the tc family of insecticidal toxins.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY220492; AAF57764.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR InterPro; IPR006530; YD.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR TIGRfams; TIGR01643; YD_repeat_2x; 1.
KW Plasmid.
SQ SEQUENCE 1433 AA; 157723 MW; 15FEDEF15118A4B CRC64;

Query Match 50.4%; Score 3984.5; DB 2; Length 1433;
Best Local Similarity 52.8%; Pred. No. 4.3e-245;
Matches 786; Conservative 187; Mismatches 432; Indels 85; Gaps 15;

Qy 1 MQNSQPSITELSLPGGGGAIITGGEALITPGDGHMAALSILPLISAGRGYAPFTLYN 60
Db 1 MQNRQDMAITTPSLPGGGAVTGLKGDIAAGPGAATLSIPLVSPGRTPTGTLYNH 60
Qy 61 SGAGNSPGLGMDCNVMTIRRTHFVPHYDETDTLFGPEGEVLVA-----DQPDSES 114
Db 61 SRAGNTFGTGLSGGPAVQVRRTRSGAPAYDETFDTPDGEVLVPAITAGTPTETQAT 120
Qy 115 TLQINLGATFTVTGRSLRSHFSRLLEYQPKTKTGTDFWLIYSPGQVHLLGKSPQAR 174
Db 121 TLGIFSGNFSQVYRSRTGSLRSLRWVPADEAETDFWLYTTPGQVALLGRNAQAR 180
Qy 175 ISNPSQTTQAMLEASVSSRGEIYYQYRABDDTCCRADEITHHLOATAQRYLHIVY 234
Db 181 ISNPARPTQAVMLMESSVSLTGEQMYQYRABDDTCCDEMRDAHPQAGQRYLVAVY 240
Qy 235 GNRTASETLPLDGSAPSQADWLFYLVFDYDYSRNNLKTTPAFSTGCS--WLCHQDRFSR 292
Db 241 GNQARALPAL-MSTPSMDSWLPVQVFDYDYSRSTALSEPPAWQTPGSGBWRRCQDCFSE 299
```


Query Match		42.4%;	Score 3350.5;	DB 2;	Length 1401;
Best Local Similarity		48.6%;	Pred. No. 1.3e-204;		
Matches 685;		Conservative 197;	Mismatches 474;	Indels 53;	Gaps 21;
QY	1	MQNSQDSITELSLPKGGGAIQMGALITGPDGMAALSPLPISAGRYAPAFITLNYN	60		
DB	1	MQGSTPLKLEIPSLPGGGSLKMGGEALNAVAGAGGASPSLPISVGRGLVPVLSLNY	60		
QY	61	SGAGNSPFGLCWNCVNTIRRTTHFGVPHYDETDFTLGPGEVL--VVAD---QP--RDES	114		
DB	61	STAGNSFGMGWCGVGFISLRTAKGVPHYTGQDEVLGPDGEVLSIVPDSQGQPEQRTAT	120		
QY	115	TLOGINLGAFTVTGYSRLESFHSRLEYWQP--KTTGKTDFWLITYSPDCOVHLLGKSPQ	172		
DB	121	SLIGTVLTQPHVTTRVQSRVAEKIVLEHWQPOORREETSFWLFTADGLVHLFGKHH	180		
QY	173	ARISNPSQTTQTAQWLLEASVSRGEQIYYQYRAEDDTGCEADEITHH--LQATAQRYLH	230		
DB	181	ARIADPODETRIARLWMEETVTHGEHIYYHYRAEDDLDCDEHELAQHSVLRPTLSWQ--	238		
QY	231	IYVYGNRTASBETPLDGSAPSQADWLFPYLVFDYGGERSNMLKTPPAPSTTGS-----	282		
DB	239	-VOYGNTQPTAFVAKSGIPVDNDLFLVDPYGERLSLNSVPFNVSENNVSNVS	297		
QY	283	--WLCRODRFSRYEYGEIRTRLCROVLMYHLQALDSKITTEHNGPTLVSLRLINYNDES	340		
DB	298	EKWRCPDSFSRYEYGEIRTRLCROVLMYHLQALDSKITTEHNGPTLVSLRLINYNDES	357		
QY	341	AIASLTIVFVRVHQEGQNVVTLPPLELAYQDFSRRHAAHWQMDVLNANFQIQRWLVD	400		
DB	358	NKVSLLQTARSLAHEITDGTVPWMSPLEMDYQVRNHGVNLNMQSPQLEKNTLQPYQLVD	417		
QY	401	LKGEPLGLLYQD-KGAWWTRSAQR-LGEIGSDAVTWKMQPLSVIPSLQSNASLVVDING	458		
DB	418	LYGEGIPALLYQDTQKAWMYRAPVRDITABGTNAVITYEEAKPLPHIPAQGESAMLLDING	477		
QY	459	DGOLDWITGPGRLGYSQRDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIG	518		
DB	478	DGRLDWITASLGLYHMGPEGBWTFPIFUSAVPMFIFHPQAKLADIDAGLPDLALIG	537		
QY	519	PKSVRLYANTRDGPAGKQVVGSDITLPPGADPRKLVAFSDVLGSGQAHLEVSATKV	578		
DB	538	PNSVRVMSNNPAGWDRAQDVHLSNKLPLPGKHKRLHVAFSDMTGGSGSHLVEVTANSV	597		
QY	579	TCFNLGRGFGQBITLPGESQAPTEFNPAQVYLADLDGSGPTLLIYVHTNRDLIFLKS	638		
DB	598	RYWPNLGHGKFGPELMTTGFQITGKRLTPTDCIWTYMAQAPDFIYARNTYLELYANES	657		
QY	639	GNGFAEPVTLRFPBGLRFDHTCOLQADVOGLGVASLILSVPHMSPHHWRCDLTNMKPWL	698		
DB	658	GNHSAEPQRIPLDGVAFRFDTCRIQIADTQGLGTASLIITIPMKVQHWRLDMTIFKPLW	717		
QY	699	LNEMNNMGVHHTLYRSSSQFWMDEKAAALTTGTQTPVCYLPFPITLWQTEDEISGN	758		
DB	718	LNAVNNMGTEITLYYSSAQFWMDEKQASESGMTVVSYLPPFVHVMRTEVLDEISGN	777		
QY	759	KLVTTLRYAGAMDGRERFRFGYVQTSHQLA---QGNAPERTPPALTQWYATGLP	815		
DB	778	RLTSHYHYSHGAWDGLRERFRGVRVQTDTIDRSASATQGTTHAEPAPSRTPVNYGTGVR	837		
QY	816	VIDNALSTYWR--DQAPAGFSRFTTQDNK--DVPLTTPEDDNRVYFNRAKGQLLRS	872		
DB	838	EVDILLFTETWQGGQQAFFPHFTFRFTYDEKSGDGMTVTP-SEGEFWLHRAKGQLLRS	896		
QY	873	ELYGLDSTNKHVYTVTEFRSQVRLRQHTDSRYFVLWSSVYESRNYHYERIASDPCSQ	932		
DB	897	ELYGDDSDILAGTFYSVDESRTQVRLPLVMVSDVPAVLVSARESQRVYERVAITDPCSQ	956		
QY	933	NITLSSDRFQOPLKQLSVQVPRROQPAINLYPDTLPKLLANSYDDOORQLRLTYQSSW	992		
DB	957	KIVLKSALGPPQDNLEIAYSRRPQPFSPYPTLPTLTSSPDEQMFRLRTRQFSY	1016		

Query Match 28.8%; Score 2277.5; DB 2; Length 1447;

Best Local Similarity 36.1%; Pred. No. 5.2e-136;

Matches 544; Conservative 225; Mismatches 595; Indels 143; Gaps 37;

QY	993	HHLT--NNTVRVLGLPDSRSDIFTTYGAENVVAGLNLLELLS--DKNSLIADDKPREYLGQ	1049		
DB	1017	HHLNHDNDNTWITGLMDTSSDARIYQADKVDPDGGFSLWFSAATGAGALLPDAADYLGH	1076		
QY	1050	QKTAYTDGQNTPLQPTTQOALIAFTTETTVFNQSTLSAFNGSPSKLSTLLEQAGYQQT	1109		
DB	1077	QRVAYTG-----PEQPAIPLPVAYTETAEPDERSLAAPFEVMDQEQLTKQLDAGWNTA	1131		
QY	1110	NYLFPRTGDKVVAHHGYTDYGTAAQFWRPKQSNQTLTKITLLIWDANYCVVUTRDA	1169		
DB	1132	KVPFSKTFHVVWVGOKETETAGADGFRPLVQRETKUTKTKTTVTDWDSHYCVITATEDA	1191		
QY	1170	AGLTTSAKTDWRFPLTPVQLTIDINDOHLITLDALGRPITLRFMTGTENGKMTGYSSEKAS	1229		
DB	1192	AGLEMQAHYDREFWADVNTDINDNVHTVTFDALGRVTSFRFWGTENGKQGYTPAENET	1251		
QY	1230	--FSPSPSDVNAIELKPLPVAQCQVYAPESWMPVLVSQKTFNRLAEQDWQKLYNARIITE	1287		
DB	1252	VPIVPTTVDLALAKPGIPVAGLVYAPLSWM--VQASFSDGDEL-YGELKXPAGIITE	1307		
QY	1288	DGRICCTLAYRRVQSQ--KAIPOLISLANNPRLPPHSLTLTTDRYDHDPEQOIRQOVVF	1345		
DB	1308	DGYLLSLAFRRHQNPPAAMPQVNSQN-----PPHVLVSITDRYDADPEQOIRQOITTF	1362		
QY	1346	SDGFGRLQLQAAARHEAGMARQNRNEDGSLI	1374		
DB	1363	SDGFGRLTQAVRHESGEAWRDEYGAIV	1391		

RESULT 14

Q4ZP56_PSES5

ID Q4ZP56_PSES5 PRELIMINARY; PRT; 1447 AA.

AC Q4ZP56_2005 (T-EMBLrel. 31, Created)

DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)

DE Salmonella virulence plasmid 65kDa B protein.

GN ORFNames=PeVr 4036;

OS Pseudomonas syringae pv. syringae B728a.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=205918;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RG DOE Joint Genome Institute;

RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,

RA Nolan M., Goltzman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,

RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.,

RT "Comparison of two complete genome sequences of Pseudomonas syringae

RT pv. syringae B728a and pv. tomato DC3000.";

RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RA Loper J.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RN [3]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RA Feil H., Feil W.S., Lindow S.E.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL: CP000075; AY39066.1; -; Genomic_DNA.

DR InterPro: IPR004048; Reg_chir_condens.

DR InterPro: IPR003284; Sal_SpVb.

DR Pfam: PF03534; SpvB; 1.

DR PRINTS: PR01341; SALSVPBPROT.

DR PROSITE: PS00626; RCCL 2; UNKNOWN 1.

SQ SEQUENCE 1447 AA; 161914 MW; B76B60FB17F29009 CRC64;

80	SEQUENCE	1446 AA; 162070 MW; 8DBAAD65DCC8949 CRC64;
	Query Match	28.8%; Score 2274; DB 2: Length 1446;
	Best Local Similarity	35.7%; Pred. No. 8.7e-136;
	Matches 550; Conservative 220; Mismatches 562; Indels 188; Gaps 39;	
Qy	2	QNSQDFITSLPSLPKGGGAIITGMEALTPGPGDMAALSPLPISAGRGYAPAFITLNYNS 61
Db	13	QSLQPAVATPL-LPKGGGAIQSIGKMGVSGTSGAASLEVALPISPGRGYAPALSLSYQS 71
Qy	62	GAGNSPGLGNDGNMTIRRTTHGPHYDETFGLGPEGEVLVADQPRDE----- 113
Db	72	TSNGVGLGWNLTNSKVARASRGVPTYANDDLIFGPGGDVCL---PERDSDGALVSTQ 128
Qy	114	-STLQGINLGAFTVTGYRSLRSHSRLEVMQPKTKGTDWMLIYSPDGOVHLGKSPQ 172
Db	129	VSRYNHDLNATVQVRYRSEVGAFLRIEHW-VDIADPGFWLHGADGSLNLYGRTS 187
Qy	173	ARISNPQTTQTQAWLLEASVSRGEQIYYQYRAEDDTGCEADBITHLQATQRYLHIV 232
Db	188	SRIADPADMRVAEWLDDSMNAIGEHLIYKPEDHQGLAED---HPRNFRAQRYLSRV 244
Qy	233	YGNRTASETL-----PGLDGSAPSQADWLFYLVFDYGRSNNLKTTPAFSTTGSMLCRQ 287
Db	245	RYGNKAKHYVYLWQEDSLDGLL-----WHFDLIIFYDQDTRSEPPPEYDEQFTWVRS 299
Qy	288	DRFSRYEYGEIRTRLCROVLMVHLQALDLSKITEHNGPTLVSRLLTYNDESAIATLV 347
Db	300	DPHSSFAYPELGNLRCROVLMPHH---FPNELGE---APLLTRRLLEHLQTLGYNLL 354
Qy	348	FVRVVGHEQ--DG-----NVVTLPLELAYQDFPRHHAHQPMDFLANFNAIORWLQVLDL 401
Db	355	---SAASHQAWDGTDRWRVQPPVQFYTDPS-LESGIYTPLEPMAGLNDGQYQLVDL 410
Qy	402	KBGLPGLLYODKGAWYRSAQRIGEIGSDAVTWKMQPLSVIPSLQS-----NASLVGIN 457
Db	411	YGDGLPGILYRDDKAMLYREPVRDITTCGADAVAYGACQPLPRIPTADSATFVRQTLTDLT 470
Qy	458	GDGOLDWITGPGLRGYSQRPDGSTRPTPLNALPVEYTHPRAQLADLACAGLSLAVLI 517
Db	471	GDGRLDWVVAQPGMAGPFTLNPORSWSNYATFSAFFAFHPQGMADLVGDGLSLDALI 530
Qy	518	GPXSRLVLYANTR-DGFAKGDVQVQSGDITLFPVGADPRKLVAFSDVLGSGQAHLEVSAT 576
Db	531	GPXSRLVLYANRRADGFAAATDIPHDED-RUPLLLSDSSTELVAFSDLLGTGQQLIRIRHN 589
Qy	577	KVTCWPNLGRGRFQCP---ITLPGFSQPAFENPAQVYLADLDGSGPTDLIYVHTNRLDI 633
Db	590	EIRVWPNLGRGRFGKGLFATLP-YTYEA--FDSSQVRLADLDGSGASDVLYLQADGFQV 646
Qy	634	FLNKSNGPAEPVTLRPEGLRPDHTCOLQWADVQGLGVASLILSVPHMSPHWH-----RC 689
Db	647	FMNRGGNGLAAPPQORWPEGVRYDRFCQSAVDLLGLGFSLSLVTVPHMAPLHWSLYAA 706
Qy	690	DLTN--MKPMLLNEMNNMNVHHTLRYRSSSQFWMDEKAAALTTGQTPVCVLPPIHTLW 747
Db	707	DRAGSVKPYLLKASDNNLGNAGEVYSRSAQEWLDEKNELRAAGSVAVSELPPVHVVV 766
Qy	748	QTEDEISGNKLVTTILYARGAWDGRERFRGFGYVEQTDSSHOLAQGNAPERTPPALTK 807
Db	767	QRTMQDKVTGNTLTQLFRYRQGFYDPRERFRGFGLLQTDTSQ-QDEDFAPVLNK 825
Qy	808	NWYATG-----LPIVDNALSTEYWDQDQAFGFSRFTTWQ 843
Db	826	TWFTGTRYPARCPTHDRSDLLARLPGEHVLSRLDAATAQTEQ----- 867
Qy	844	DNKDVLPTPEDDNRVWFNRALKQLLRSELYGLDDSTNKHVPYTVTFRSQVRRLQHTD 903
Db	868	-----PITDADATLQEMARALSGSVLRSEVFGLDASQRPVLYSTRSCRVLVRLQALS 922
Qy	904	SRYVLWSSV-----VESRNYHE-RIASDPOCSQNTILSSDRFQPLKQLSVQYPRRQOP 958
Db	923	AHRP--YASMLPLSLVITYEAELEEDPMCEHSLNLAWDRYGSLTSHSVSNYARRKCP 980

Qy	959	A-INLYPDTLPKLLANSYDDQORQLRITYQOSSMHHLTNNTVRVLGLPDSTRSDIFTYG 1017
Db	981	GUAPPADPHQQQWEASHDDAQOQPYLNMHAEALYLDSPQSWRLGLPVRTGDAWLIP 1040
Qy	1018	AENVPAAGNLLELLSKNSLIADDKPREYLGGQKTAYT---DGQNTTPIQTPTTROALIAF 1074
Db	1041	ASALTPAQISYEQFADPSGFFA-TLPRTLTSLSVQRYIGCGDE-----ASFQALADA 1092
Qy	1075	TETTVNQSTLSAFNGSIPSDKLSLTLEQAGVQOQNYLFPRTGEDKVMVAHHCYTDYGA 1134
Db	1093	VETAELEDDHALSAYERVMDSVTLAEKLEIGYQQMPSFLPADSLN-LMSVKRGFATYAGQ 1151
Qy	1135	AQWRPQKQSNTOCTGKITLIWDANCVVVQTRDAAGLTTSKAYDWRFLTPVOLTDINDN 1194
Db	1152	EHPFHTQTPRFRSHGWSLVEYDAYHLFATRIDPADCVTTAEYDYRVLPQKRIIDPNQN 1211
Qy	1195	QHLITLDALGRPTTLRPMGTENGKMTGYSPKASFPSPDVNAALIELKKPLPVAOCQVY 1254
Db	1212	QREADYDAFGRVWATSFYGTGELGEAVGPPPLNRAGHYWASAGEVALQ----- 1258
Qy	1255	APESWMPVLSQKTPNPLAESQWKLYNARIITEDGRICTLAYRRWTSQKAIPOQLISLLN 1314
Db	1259	-PE-----YALGRQ-----ASALYYDGN-----TVLGLVH 1382
Qy	1315	NGPRLPPHSLTLTTDRYDHDPEOQIRQQVVFSDGFGRLQAAARHEAGMARQNRNEDGSLI 1374
Db	1283	-----IPLATAVLVADRYPEDLDKQIRISMASIDGFORLTQTRKVEDGDAYSVDWGNLE 1338
Qy	1375	I-----NVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWMRYVNSDSARQEKEAYADTH 1429
Db	1339	LVDGPKIVHASPRWRISERVEYNNKGLAVRVYRPFYFANSHLYVNDASIRSQN--IVDKQ 1396
Qy	1430	VYDPIGREIKVITAKGFRITLTPMPTVNEDENDTAAEV 1469
Db	1397	FYDPLGRPTTITAKGMRRRQTYRVWYTTISEDENDTABEV 1436

Search completed: December 16, 2005, 14:41:34
Job time : 281 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:25:20 ; Search time 211 Seconds
(without alignments)
3069.404 Million cell updates/sec

Title: US-10-706-424-10
Perfect score: 7901
Sequence: 1 MQNSQDSITELSLPKGGGA.....WFTVNEDENTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003s.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7901	100.0	1474	8	Adp18622 Photorhab
2	7901	100.0	1474	8	Adr21525 Photorhab
3	7285.5	92.2	1476	6	Abm70222 Photorhab
4	6085.5	76.8	1477	6	Abm70230 Photorhab
5	6042	76.5	1481	2	Aay33730 Photorhab
6	6031.5	76.3	1476	5	Abg32852 P. lumine
7	6025.5	76.3	1476	8	Adr21576 Photorhab
8	4627.5	58.6	1486	6	Abm70526 Photorhab
9	4595.5	58.2	1485	8	Adr21578 Photorhab
10	4581.5	58.0	1485	2	Aaw56546 Toxin Tca
11	4580.5	58.0	1485	2	Aaw17887 Photorhab
12	4261	53.9	1506	8	Adr21529 Xenorhabd
13	4261	53.9	1506	9	Aeb47812 Native Xp
14	3971	50.3	1428	4	Aay97695 Sepb prot
15	3971	50.3	1428	8	Adr21540 Serratia
16	3834	48.5	1493	8	Adr20369 Recombina
17	3834	48.5	1493	8	Adr21498 Xenorhabd
18	3106.5	39.3	1444	8	Adm61347 Paenibaci
19	3106.5	39.3	1444	8	Adr21520 Paenibaci
20	1744.5	22.1	662	3	Aay95685 Cosmid cH
21	1262	16.0	697	8	Adm61375 Paenibaci
22	839	10.6	591	2	Aar23006 Protein t
23	199	2.5	2060	3	AAE20967 Staphyloc
24	172.5	2.2	1404	7	Adc00960 Enterohae

25	170	2.2	593	6	AAO19846
26	169.5	2.1	14130	9	AEA07438
27	162.5	2.1	16990	9	AEA07436
28	161	2.0	1863	7	ADD46989
29	161	2.0	1863	9	ADX26456
30	160.5	2.0	1385	6	ABU40318
31	160.5	2.0	1426	3	ABU15983
32	160.5	2.0	1426	6	ABU14693
33	159.5	2.0	1329	4	ABG28402
34	159.5	2.0	1329	4	ABG25780
35	159.5	2.0	1329	7	ADG33172
36	155.5	2.0	658	7	ABO76164
37	155.5	2.0	1400	7	ADCO1365
38	155.5	2.0	2315	7	ADFO5149
39	154	1.9	4630	3	AA777177
40	154	1.9	5215	8	ADL91930
41	153.5	1.9	1028	7	ADF06896
42	152.5	1.9	1669	8	ADU07870
43	152	1.9	4630	2	AAW19629
44	151.5	1.9	1904	5	ABB57100
45	151.5	1.9	1904	9	ADX26385

ALIGNMENTS

RESULT 1

ID	ADP18622	standard; protein; 1474 AA.
XX	XX	
AC	ADP18622;	
XX	XX	
DT	12-AUG-2004	(first entry)
XX	XX	
DE	DE	Photorhabdus luminescens W-14 tcdB2 protein SEQ ID NO:10.
XX	XX	
KW	KW	tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;
KW	KW	orally active insect toxin; insect toxin; tcdB2.
XX	XX	
OS	OS	Photorhabdus luminescens.
XX	XX	
PN	PN	WO2004044217-A2.
XX	XX	
FD	FD	27-MAY-2004.
XX	XX	
PF	PF	12-NOV-2003; 2003WO-IB005553.
XX	XX	
PR	PR	12-NOV-2002; 2002US-0425672P.
XX	XX	
FA	FA	(UYBA-) UNIV BATH.
XX	XX	
PI	PI	Ffrench-Constant RH, Waterfield NR;
XX	XX	
DR	DR	WPI; 2004-411735/38.
DR	DR	N-PSDB; ADP18621.
XX	XX	
PT	PT	New isolated Photorhabdus luminescens nucleic acids, useful for
PT	PT	expressing orally active insect toxin or for generating transgenic plants
PT	PT	with enhanced resistance to insects.
XX	XX	
FS	FS	Claim 1; SEQ ID NO 10; 118pp; English.
XX	XX	

The present invention describes DNA sequences from the tcd genomic region of Photorhabdus luminescens W-14. Also described: (1) a transgenic monocot or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of Photorhabdus luminescens W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to

CC	insects. The present sequence represents Photorehabdus luminescens W-14									
CC	tcdB2, which is used in the exemplification of the present invention.									
XX										
SQ	Sequence 1474 AA;									
	Query Match	100.0%;	Score	7901;	DB	8;	Length	1474;		
	Best Local Similarity	100.0%;	Pred. No.	0;						
	Matches	1474;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MONSQPSITELSLPKGGGAI	TGMEAL	TPGDDGMAALS	SLPLPIS	SAGRGY	APAF	TLN	YN	60
DB	1	MONSQPSITELSLPKGGGAI	TGMEAL	TPGDDGMAALS	SLPLPIS	SAGRGY	APAF	TLN	YN	60
QY	61	SGAGNSPGLGWCN	VMTIRRR	THFGVPHYDE	TD	PLGPEGE	VLVADQ	PRDE	STLQ	GIN
DB	61	SGAGNSPGLGWCN	VMTIRRR	THFGVPHYDE	TD	PLGPEGE	VLVADQ	PRDE	STLQ	GIN
QY	121	LGATFTVGYRSL	SHFSRL	EWQPKTK	TKTDF	FWLIYSP	DGQVHLLG	KSPQ	ARISN	PSQ
DB	121	LGATFTVGYRSL	SHFSRL	EWQPKTK	TKTDF	FWLIYSP	DGQVHLLG	KSPQ	ARISN	PSQ
QY	181	TTOTAQWLL	EA	SVSSRGE	QIYYQ	YRAEDD	TGCE	AD	EIT	THLQ
DB	181	TTOTAQWLL	EA	SVSSRGE	QIYYQ	YRAEDD	TGCE	AD	EIT	THLQ
QY	241	ETLPGLDGS	APSO	ADWLFY	LVFDY	GERSN	NKTP	PAFST	TG	SWLC
DB	241	ETLPGLDGS	APSO	ADWLFY	LVFDY	GERSN	NKTP	PAFST	TG	SWLC
QY	301	TRBLCRQV	LMHHLQ	ALD	SKITE	HNGPT	LV	SR	LIT	YDE
DB	301	TRBLCRQV	LMHHLQ	ALD	SKITE	HNGPT	LV	SR	LIT	YDE
QY	361	VTLPPL	ELAYQ	DFSP	PRHAWQ	PM	DVLAN	FNA	TOR	WOL
DB	361	VTLPPL	ELAYQ	DFSP	PRHAWQ	PM	DVLAN	FNA	TOR	WOL
QY	421	SAORL	GIGSD	AVT	WEKQ	PLSV	TPSI	QSN	AS	LV
DB	421	SAORL	GIGSD	AVT	WEKQ	PLSV	TPSI	QSN	AS	LV
QY	481	GSWTR	FTPL	NALP	VEY	THP	RAQ	LA	DL	MG
DB	481	GSWTR	FTPL	NALP	VEY	THP	RAQ	LA	DL	MG
QY	541	SGDIT	LP	VP	GAD	PR	KLV	AF	SD	VL
DB	541	SGDIT	LP	VP	GAD	PR	KLV	AF	SD	VL
QY	601	PATF	NP	QA	QV	YL	AD	LD	GS	GP
DB	601	PATF	NP	QA	QV	YL	AD	LD	GS	GP
QY	661	QLQ	MA	DV	QGL	GV	AS	LL	IS	VP
DB	661	QLQ	MA	DV	QGL	GV	AS	LL	IS	VP
QY	721	WLDE	KAA	AL	TG	QTP	VC	YL	PP	HT
DB	721	WLDE	KAA	AL	TG	QTP	VC	YL	PP	HT
QY	781	FGV	VE	Q	T	D	S	H	Q	L
DB	781	FGV	VE	Q	T	D	S	H	Q	L
QY	841	TWQ	N	K	D	V	P	L	T	P
DB	841	TWQ	N	K	D	V	P	L	T	P
QY	901	HTD	S	R	Y	P	V	L	M	S
DB	901	HTD	S	R	Y	P	V	L	M	S

RESULT 2
ADR21525
ID ADR21525 standard; protein; 1474 AA.
XX ADR21525;
AC AC
XX XX
DT 04-NOV-2004 (first entry)
DE Photorehabdus TcdB2 toxin SEQ ID NO:45.
XX toxin; insect; insecticidal; transgenic; pest control.
XX Photorehabdus luminescens.
XX WO2004067727-A2.
XX 12-AUG-2004.
XX 07-JAN-2004; 2004WO-US000394.
XX 21-JAN-2003; 2003US-0441723P.
XX (DOWC) DOW AGROSCIENCES LLC.
XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
XX DR N-PSDB; ADR21524.
XX Controlling or inhibiting an insect, useful for pest control, comprises
XX contacting the insect with effective amounts of a Protein A, a Protein B,
XX and a Protein C.
XX Claim 1; SEQ ID NO 45; 368pp; English.

CC The invention relates to a novel method for controlling or inhibiting an
 CC insect comprising contacting the insect with effective amounts of a
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
 CC is encoded by a naturally occurring gene or has an amino acid sequence
 CC that differs from the product encoded by a naturally occurring gene only
 CC by truncation or by conservative amino acid changes. Protein A is a 230-
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic
 CC species, has stand alone insecticidal activity, and has an amino acid
 CC sequence at least 40% identical to a sequence selected from XpA2w1,
 CC XpA2w1, TcdA, TcdA2, and TcdA. Protein B is a 130-180 kDa toxin
 CC complex potentiator having an amino acid sequence at least 40% identical
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XpC1w1, XpC1Bxb,
 CC PptB1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
 CC having an amino acid sequence at least 35% identical to a sequence
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XpC1w1, XpC1Bxb, PptC1
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein
 CC C. The method is useful for pest control. The present sequence represents
 CC Photorhabdus luminescens TcdB2 toxin.

XX
 SQ Sequence 1474 AA;

Query Match 100.0%; Score 7901; DB 8; Length 1474;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSQDSFISITSLPKGGGAIATGGEALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60
 Db 1 MNSQDSFISITSLPKGGGAIATGGEALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60

Qy 61 SGAGNSPFLGWCNVMIRRTTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
 Db 61 SGAGNSPFLGWCNVMIRRTTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120

Qy 121 LGATFTVTGYSRLESFHSRLEYQPKTTGKTDFWLIYSPDGQVHLLGKSPQARISNPSQ 180
 Db 121 LGATFTVTGYSRLESFHSRLEYQPKTTGKTDFWLIYSPDGQVHLLGKSPQARISNPSQ 180

Qy 181 TTQTAQMLLEASVSRGEIYYQYRAEDDTGCEADEITHLQATQRYLHIVYVYGNRTAS 240
 Db 181 TTQTAQMLLEASVSRGEIYYQYRAEDDTGCEADEITHLQATQRYLHIVYVYGNRTAS 240

Qy 241 ETLPLDGSAPSQADWLFYLPDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEGFEIR 300
 Db 241 ETLPLDGSAPSQADWLFYLPDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEGFEIR 300

Qy 301 TRRLCROVLMYHLQALDSKITEHNGPTLVSRLLINLYNDESAIASTLVFVRVGHQDGNV 360
 Db 301 TRRLCROVLMYHLQALDSKITEHNGPTLVSRLLINLYNDESAIASTLVFVRVGHQDGNV 360

Qy 361 VTLPPLLEAYQDFSPRHHAHQPMQDVLNFAINAIQWQLVDLKGELPGLLYQDRGAWMYR 420
 Db 361 VTLPPLLEAYQDFSPRHHAHQPMQDVLNFAINAIQWQLVDLKGELPGLLYQDRGAWMYR 420

Qy 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDGLDWITTPGLRGRVHSORPD 480
 Db 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDGLDWITTPGLRGRVHSORPD 480

Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRDGFAGKQVQV 540
 Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRDGFAGKQVQV 540

Qy 541 SGDITLPPVGADPRKLNAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQIPITLPGFSQ 600
 Db 541 SGDITLPPVGADPRKLNAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQIPITLPGFSQ 600

Qy 601 PATEFNPAQVYLAIDLGGSGPTDLIVYVHTNRDLIDIFLNKSGNGFASBPVTLRPFEGRLRPH 660
 Db 601 PATEFNPAQVYLAIDLGGSGPTDLIVYVHTNRDLIDIFLNKSGNGFASBPVTLRPFEGRLRPH 660

Qy 661 QLOMADVOGLGVASLIISVPHMSPHHWCDITNNKPKLLNEMNNMNGVHHTLYRSSSQF 720
 Db 661 QLOMADVOGLGVASLIISVPHMSPHHWCDITNNKPKLLNEMNNMNGVHHTLYRSSSQF 720

Qy 721 WLDEKAAALTTGQTPVCYLPPFIHTLMQTEDEISGNKLVTTILRYARGAWDGREREPRG 780
 Db 721 WLDEKAAALTTGQTPVCYLPPFIHTLMQTEDEISGNKLVTTILRYARGAWDGREREPRG 780

Qy 781 FGYYEQTDSHQLAGNAPERTPPALTGNWATGLPVDNALSTEVWRDDQAFAGFSRPT 840
 Db 781 FGYYEQTDSHQLAGNAPERTPPALTGNWATGLPVDNALSTEVWRDDQAFAGFSRPT 840

Qy 841 TWQNDKDVLTPTDDNSRYFNFRALKQGLLRSELYGLDDSTNKHVPYTVTFEFSQVRLQ 900
 Db 841 TWQNDKDVLTPTDDNSRYFNFRALKQGLLRSELYGLDDSTNKHVPYTVTFEFSQVRLQ 900

Qy 901 HTDSRYPLVSSVSVESRYHYERIASDPSQNTLSSDRFGQPLKOLSVQYPRQOPAI 960
 Db 901 HTDSRYPLVSSVSVESRYHYERIASDPSQNTLSSDRFGQPLKOLSVQYPRQOPAI 960

Qy 961 NLVYPTLTPDKLLANSYDDQORQLRLTYQQSSWHHLTNTVVRVLGLPDSRSDIFTYGAEN 1020
 Db 961 NLVYPTLTPDKLLANSYDDQORQLRLTYQQSSWHHLTNTVVRVLGLPDSRSDIFTYGAEN 1020

Qy 1021 VPAGLNLLELLSKNSLIADDKPREYLGQOKTAYTDGQNTTLPQTPTQALIAFTETV 1080
 Db 1021 VPAGLNLLELLSKNSLIADDKPREYLGQOKTAYTDGQNTTLPQTPTQALIAFTETV 1080

Qy 1081 NQSTLSAFNGSIPSDKLSITLLEQAGYQOTNYLPRTGEDKVVVAHGYDYGTAQFWRP 1140
 Db 1081 NQSTLSAFNGSIPSDKLSITLLEQAGYQOTNYLPRTGEDKVVVAHGYDYGTAQFWRP 1140

Qy 1141 QKQNTOLTKKITLWDANYCVVVQTRDAAGLTTSKVDWRFLLTPVOLTDINDNQHLLITL 1200
 Db 1141 QKQNTOLTKKITLWDANYCVVVQTRDAAGLTTSKVDWRFLLTPVOLTDINDNQHLLITL 1200

Qy 1201 DALCRPTTLFAFWGTENGKMTGYSPEKASFPSPSDVNAALTELKPLPVAQCQVYAPSSWM 1260
 Db 1201 DALCRPTTLFAFWGTENGKMTGYSPEKASFPSPSDVNAALTELKPLPVAQCQVYAPSSWM 1260

Qy 1261 PVLQKTFNRLAEOQWOKLYNARIITEDGRICITLAYRRVWQSKAIIPOLISLLNNGPRLP 1320
 Db 1261 PVLQKTFNRLAEOQWOKLYNARIITEDGRICITLAYRRVWQSKAIIPOLISLLNNGPRLP 1320

Qy 1321 PHSITLTTDDYDHDPEQIIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLIINVQHT 1380
 Db 1321 PHSITLTTDDYDHDPEQIIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLIINVQHT 1380

Qy 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSAROEKAYADTHVVDPIGREIKV 1440
 Db 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSAROEKAYADTHVVDPIGREIKV 1440

Qy 1441 ITAGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474
 Db 1441 ITAGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474

RESULT 3
 ABM70222
 ID ABM70222 standard; protein; 1476 AA.
 XX AC ABM70222;
 XX DT 20-NOV-2003 (first entry)
 XX DE Photorhabdus luminescens protein sequence #3319.
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX whoping cough.
 OS Photorhabdus luminescens.
 XX WO200294867-A2.
 XX

RESULT 4

ABM70230
ID ABM70230 standard; protein; 1477 AA.

XX AC ABM70230;

XX XX 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #3327.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN W0200294867-A2.

XX XX 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX XX 07-FEB-2001; 2001FR-00001659.

XX PR (INSP) INST. PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Prangeul L, Kunst F, Danchin A;
XX PI Buchrieser C;

XX DR WPI; 2003-148459/14.

XX XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 3327; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded
XX CC proteins from Photorhabdus luminescens. The isolated sequences are
XX CC sources of probes and primers for detecting the genome of P. luminescens
XX CC and related species; to study polymorphisms; for gene analysis and for
XX CC detection/amplification of the genes. Antibodies (Ab) raised against the
XX CC polypeptides encoded by the genes are used for detection/identification
XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX CC carry a gene-containing vector are used to select compounds that
XX CC modulate, regulate, induce or inhibit expression of the genes in plants,
XX CC animals or microorganisms other than P. luminescens and are able to alter
XX CC response or sensitivity to toxins and antibiotics produced by P.
XX CC luminescens. Cells transformed to express the genes are useful for
XX CC recombinant production of the proteins, particularly toxins and
XX CC antibacterials useful as insecticides, bactericides and fungicides. The
XX CC genes, proteins, vectors containing the genes and Ab are also useful
XX CC therapeutically (to treat microbial infection by bacteria or fungi that
XX CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX CC biopesticides. Other uses of the genes and the proteins are as virulence
XX CC factors and for identifying targets of human diseases for which P.
XX CC luminescens is a model (particularly plague and whooping cough). This
XX CC sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 1477 AA;

Query Match 76.8%; Score 6065.5; DB 6; Length 1477;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

QY 1 MNSQDFSIITSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGYAPFTLNYN 60

DB 1 MNSQTFVSELSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGYSPSLNYS 60

QY 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

DB 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVAIALNENGQADISES 120
QY 115 TLOGINLGATFTVGYRSRLESFHSRLEYWQPKTKTDFWLIYSPDQVHLLGSKSPQAR 174
DB 121 SLQGINLGEFTVGYRSRLESFHSRLEYWQPKTKTGTDFWLIYSPDQVHLLGSKNQAR 180
QY 175 ISNPSQTTQAWMLLEASVSSRGEQIYYQRAEDDTCCEADEITHHQAQRIHLVYY 234
DB 181 ISNPLNVSTQAWMLLEASVSSHGEQIYYQRAEDDTCCEADEITHHQAQRIHLVYY 240
QY 235 GNTASSTLPLGLDGSAPSQADWLFYLVDFYDGERSNLKTTPAPSTTSGMLCRODRFRYE 294
DB 241 GNTASEVFTLGGDDPLKSGMLFCLVDFYGERKNSLSEIPPFKASSMLCRODRFRYE 300
QY 295 YGFEIRTRLCROVLMYHHLQALDSKITEHNGTFLVSRLILNLYDESAIASTLVFVRVGH 354
DB 301 YGFELTRRLCRQLIMPHRLQTLGGQAKGDDEPALVSRLLDYDENAVISTLVSVRRIGH 360
QY 355 EQDGNVVTLPPLLEYAYQDFSPRHHAHWPMQDVLANFNAIORWQLVDLKGELPGLLYQDK 414
DB 361 EDNNTVISLPLLEYAYQDFPEBQKARQMSQMDVLANFNAIORWQLVDLKGELPGLLYQDR 420
QY 415 GAWYRSQORLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWVITGPGLRGY 474
DB 421 NGWYRSQORLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWVITGPGLRGY 480
QY 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIGPKSVRLYANTRDGFAP 534
DB 481 HSQHPDGSWTRFTPLDALPIEYSHPRAQLADLMAGLSDLVLIGPKSVRLYANNRDGFQ 540
QY 535 GKDVVQSGDITLPPVGADPRKLVAFSVLSGQAHLYEVSAKVTCVTPNLRGRFGQPIT 594
DB 541 GRDVVQSGDITLPLPGADARKLVAFSVLSGQAHLYEVSAKVTCVTPNLRGRFGQPIT 600
QY 595 LPGFSQATFENPAQVYLADLDCGSPDLYVHTNRDLDFLNKSGNGFAEPVTLRFPEGL 654
DB 601 LPGFSQADNFNDPRVHLADLDCGSPADLYVHTNRDLDFLNKSGNGFAEPVTLRFPEGL 660
QY 655 RFDTTCQLQADVQGLGAVSLILSVPHMSPHWRCDLTNNKMPLLNMMNNMNVHHTLRY 714
DB 661 RFDDTCQLQADVQGLGAVSLILSVPHMAPHWRCDLTNNKMPLLNMMNNMNVHHTLRY 720
QY 715 RSSSQFQLDEKAAALTTGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYARGANDGR 774
DB 721 RSSSQFQLDDKAAALATGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYARGANDGR 780
QY 775 EREPRGFGYVEQTDHQAQGNAPERTPPALTKNWYATGLPVDNALSTEYWR-DDQAFA 833
DB 781 EREPRGFGYVEQTDHQAQGNAPERTPPALTKNWYATGLPVDNALSTEYWR-DDQAFA 840
QY 834 GFSPRFTTWQDNKDVPLTPDDNSRYWFRNALKGQLRSELYGLDDSTNKHVPYTVTFER 893
DB 841 GFTPRFTRWKQDVPAVPENDNLYWFRNALKGQLRSELYGLDDSEQQNIPYTVTESR 900
QY 894 SQVRLQHTDSRPVPLWSSVVSRYHYRIASDPQCSQNTLSSDRFGQPLKQLSVQYP 953
DB 901 PQVRQLQDGTASPVLSWASVVENRSHYRIIGDPQCNQDITLSSDQFGQPLKQVWQYP 960
QY 954 RROQPAINLYPDTLPDKLLANSYDDQORLRLTYQSSWHHTNTNRYVTLGLPSTRSDI 1013
DB 961 RRNPQTTNPYPDTLPDLFPASSYDDQOQLRLTCQSSWHHTNTNRYVTLGLPSTRSDA 1020
QY 1014 FTYGAENVVAGGLNLELLSDKNLSIADDKPREYLGQCKTAYTDGQNTTLPQTPTROALIA 1073
DB 1021 FTYDAKQVPDGLNLEALCHENSIIADDKPREYLNQORTFTYDQKNQAPLEIPTRQALIA 1080
QY 1074 FTETVFNQSTLSAFNGSIPSDKLSITLLEAGYQOQNTNLPFRTGEDKVVVAHHGYDTGT 1133
DB 1081 FTETAVLTESLLSAFDGGITPDELPGILTQAGYQOEPYLPFRTGENKVVVAHQYDTGT 1140
QY 1134 AAQWRKQKSNLTQTKITLILWDANYCVVQVTFDAAGLTTSYAKYDWRFLTPVQLTDIND 1193

Db 1141 EAQFWRPVQAQRNTLLTGKTTLQMDTHYCVITQTQDAAGLTVLANYDWRFLTPVQLTDIND 1200
Qy 1194 NOHLITLDALGRPTLRFWCTENGKMTGYSSPEKASFPSPDYNAATELKKPLPVAQCQV 1253
Db 1201 NVHLITLDALGRPTLRFWCTENGKMTGYSSPEKSPFPIDINTALTGPLPVAQCLV 1260
Qy 1254 YAPESWMPVLSQKTFNRLAEPQWQKLYNARIITEDGRICTLAYRRWQSQKAIPLQISLL 1313
Db 1261 YAPDSWMPVLSQKTFNRLAEPQWQKLYNARIITEDGRICTLAYRRWQSQKAIPLQISLL 1320
Qy 1314 NNGPRLPHSLTLTTRDYRHDHPQIQIRQVFSDFGRLLOAAARHAGMARQNEGSL 1373
Db 1321 TNSIGLPHSLTLTTRDYRHDHPQIQIRQVFSDFGRLLOAAARHAGMARQNEGSL 1380
Qy 1374 IINVQHTENWAVTGRTEYDNKQPIRTYQPYFLNDWRYVNSDSARKEAYADTHYDP 1433
Db 1381 VTQMEDTKTAWVGRTEYDNKQPIRTYQPYFLNDWRYVNSDSAR--KGAYADTHYDP 1438
Qy 1434 IGRKIVITAKGWRFRRLTFTPMFVNEDENDTAAEV 1469
Db 1439 IGRKIVITAKGWRFRRLTFTPMFVNEDENDTAAEV 1474

RESULT 5
AAV33730
ID AAV33730 standard; protein; 1481 AA.

AC AAV33730;
XX
XX 09-NOV-1999 (first entry)
XX Photorhabdus luminescens 1481 amino acid insecticidal toxin.
XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
XX Photorhabdus luminescens.
XX WO9942589-A2.
XX 26-AUG-1999.
XX 18-FEB-1999; 99WO-EP001015.
XX 20-FEB-1998; 98US-00027080.
XX 20-JAN-1999; 99US-0116439P.
XX (NOVS) NOVARTIS AG.
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM;
PI Chen JS;
XX WPI; 1999-527479/44.
XX DR N-PSDB; AA206831.
XX New nucleic acid from Photorhabdus luminescens encoding insecticidal
XX toxins, used for making resistant transgenic plants.
XX Claim 26; Page 140-145; 148pp; English.
XX This sequence represents a 1481 amino acid insecticidal toxin from
XX Photorhabdus luminescens. It is one of three insecticidal toxins
XX (AAV33728-Y33730) encoded by open reading frames (orfs) in a 39kb
XX fragment of P. luminescens DNA (AAZ06831). This sequence is encoded by
XX orf2. P. luminescens is a member of the Enterobacteriaceae family and is
XX a symbiotic bacterium of nematodes of the genus Heterorhabditis. The
XX nematodes colonise insect larvae, kill them, and their offspring feed on
XX the dead larvae. However, the insecticidal agents are produced by P.
XX luminescens rather than the nematodes. The toxins have activity against
XX Lepidopteran insects such as Cabbage looper (Trichoplusia ni), European
XX Corn Borer (Ostrinia nubilalis) and Fall Armyworm (Spodoptera frugiperda)
XX and also against Coleopteran insects (e.g., Colorado Potato Beetle,
XX Leptinotarsa decimlineata). In addition the toxins are active against

CC strains resistant to known insecticides. The DNA sequence can be used to
CC generate transgenic plants of various species that are resistant to
CC economically important insect pests and also for recombinant production
CC of the toxins for use as insecticides
XX
SQ Sequence 1481 AA;
Query Match 76.5%; Score 6042; DB 2; Length 1481;
Best Local Similarity 76.1%; Pred. No. 0;
Matches 1127; Conservative 113; Mismatches 226; Indels 14; Gaps 5;
Qy 1 MONSQFSPITELSLPKGGAITGMEALTPGPDGMAALSPLPISAGRGYAPFTLYN 60
Db 1 MONSQFSPITELSLPKGGAITGMEALTPGPDGMAALSPLPISAGRGYAPFTLYN 60
Qy 61 SGAGNSPFGMGWDCNVTIRRTTHFGVPHYDETDTPFGPEGEVLLVA-----DQPRDES 114
Db 61 SGTGNSPFGMGWDCNVTIRRTTHFGVPHYDETDTPFGPEGEVLLVA-----DQPRDES 120
Qy 115 TLQGINLGATFTVTGYSRLESHPFSLRYWQPKTKGTDFPMLIYSPDQVHLGKSPQAR 174
Db 121 SLOGINLGATFTVTGYSRLESHPFSLRYWQPKTKGTDFPMLIYSPDQVHLGKSPQAR 180
Qy 175 ISNPSQTTTQAOMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 234
Db 181 ISNPLVNVQTAQWMLLEASVSSHGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 240
Qy 235 GNRASSETLPGLDGSAPSQADWLFPYLVDFYGERSNMLKTPPAFTSTGSMWLCRODRFSRYE 294
Db 241 GNLTADEVPTLNGDDPLKSGWFLCLVDFYGERKNSLSEMPFPKATSNWLCRKDRFSRYE 300
Qy 295 YGFEIRTRLCROVLMYHHLQALDLSKITEHNGPTLVSRLLNYDESAIATSLVFRVRVGH 354
Db 301 YGFAIRTRLCROVLMYHHLQALDLSKITEHNGPTLVSRLLNYDESAIATSLVFRVRVGH 360
Qy 355 EQDG-NVVTLPPLLELAYQDFSPRRHAWQPMVDVLANFNATQRLQVLDLKGELPGLLYQD 413
Db 361 EQDGTAVVALPPLLELAYQDFSPRRHAWQPMVDVLANFNATQRLQVLDLKGELPGLLYQD 420
Qy 414 KGAWYRSRQRLGEISDVAVTWEMQPLSVIPSLQSNASLVDINGDQLDQWVTGPGLRG 473
Db 421 KNGWYRSRQRLGEISDVAVTWEMQPLSVIPSLQSNASLVDINGDQLDQWVTGPGLRG 480
Qy 474 YHSQRPDGSMTRFTPLNALPVEYTHPRAQLADLMGAGLSLVLIGPKSVRLYANTRDGEA 533
Db 481 YHSQHPDGSMTFTPLHALPIEYTHPRAQLADLMGAGLSLVLIGPKSVRLYANTRDGEA 540
Qy 534 KGKDVVQSGDITLTPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNPLGRGFCQPI 593
Db 541 EGRDVVQSGDITLTPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNPLGRGFCQPI 600
Qy 594 TLPFGSQPATEFNPQAVYLLADLQSGPTDLIYVHTNRDLIFLNKSGNGPAPVTLRPPG 653
Db 601 TLPFGSQSAAANFNDRVHLADLQSGPADLIYVHADHLDIFSNEGNGPAPVTLRPPG 660
Qy 654 LRFDTHTCOLQADVOGLGVASLILSVPHMSPHWRCDLTNMKPLLNMNNMNVHHTLR 713
Db 661 LRFDDTTCQLQADVOGLGVVSVLILSVPHMAPHWRCDLTNAKPLLSEMMNNMAHHTLR 720
Qy 714 YRSSSFMLDEKAAALTTGQTPVCYILPFIHTLWQTEDEISGNKLVTTLRYARGAWDG 773
Db 721 YRSSVQFMDLDEKAAALTTGQTPVCYILPFIHTLWQTEDEISGNKLVTTLRYARGAWDG 780
Qy 774 REREFRGVGVTEQDSHQLAQGNAPERTPPALTKNWTATGLPVIDNALSTLEYWR-DDQAF 832
Db 781 REREFRGVGVTEQDSHQLAQGNAPERTSPALTKNWTATGLPVIDNALSTLEYWR-DDQAF 840
Qy 833 AGSPRFTTWQDNKDVPLTPEDDNRVFNRLKGLLRSELYGLDDSTNKHVPTVTFE 892
Db 841 TGTPIHFTLWKSQDVPLTPEDDNRVFNRLKGLLRSELYGLDDSTNKHVPTVTFE 900
Qy 893 RSQVRRLQHTDSRPYPLWGSVSVSRNYHYERIASDPQCQNITLSSDRFQPLKQISVQY 952

Db 901 RPQVRLQDNNTTSPVLWASVVSRSYHYERIISDPQCNDQDITLSSDLFGQPLKQVSQY 960
QY 953 PRQOPAINLYPDTLPDKLLANSYDDQORQLRLTYQSSWHLLTNTVTVLGLDSTSD 1012
Db 961 PRRKPTTNPYPDTLPDTLPASSYDQOQLRLTYQSSWHLLTANELRLVLGLPDGTRSD 1020
QY 1013 IFTYGAENVPAAGNLLELLSKNSLIADDPREYLGQOKTAY----TDGNTTPTLQTPTR 1068
Db 1021 AFTYDAKHVPDGLNLEALCAENSLIADDPREYLNQRTYPTDGTDKGKPTPLKTPTR 1080
QY 1069 QALIAFTTETTVFNQSTLSAFNGSPSKLSTLLEQAGYQOQNTYLPFRPTGDKVWVAHGY 1128
Db 1081 QALIAFTTETAVLTSLLSAFDGGITPDELPGLLTQAGYQOQNPYLPFLSGENQVWVARKGY 1140
QY 1129 TDYGTAAQFWRPQSQNTQLTKTLIWDANYCVVQTRDAAGLTTSKAKYDWRFLTPVOL 1188
Db 1141 TDYGTGQFWRPVAQRNTQLTKTLKWDTHYCVITQTDAGLTVSANYDWRFLTPMQL 1200
QY 1189 TDINDNOHLITLDALGRPTTLRFWTGKNGTGYSSPEKASFPSPDVNAALIELKPLPV 1248
Db 1201 TDINDNVHIIITLDALGRPTVQRFNGIENGVAIGSSPEAKPFTPPVDVNAALITGLPLV 1260
QY 1249 AQCVYAPESWMPVLSQKTFNRLAEQDMQKLYNARIITEDGRICITLAYRRWVQSKAIPQ 1308
Db 1261 AQLVYAPDSWMPVLFQGTFTNTLTQEQKTLRLDRIITEDWRICALARRMLQSKAGTP 1320
QY 1309 LISLNNQPRLLPHSLTLTTRYDHDPEQOIRQOVVFSDFGRLLLOAAARHEAGMARQNR 1368
Db 1321 LVKLLTNSIGLPPHNLMLATDRYDRDSRQOIRQOVAFSDFGRLLIQAARHEAGEAWQRN 1380
QY 1369 EDGSLINQVQHTENRWAVGTEDYDNKGPIRTQVOPFLNDWRYVNSDSARQEAAYDT 1428
Db 1381 QDGLSVTKQEDTKRWAITGTREYDNKGQARTQVOPFLNDWRYVNSDARSAR--KEAYADT 1438
QY 1429 HYPDIPGREIKVITAKGFRRTLPFPWFTNEDNDTAAE 1468
Db 1439 HIYDIPGREIQVITAKGWLQRNQVFPWFTVSEDNLSAD 1478
RESULT 6
ID ABG32652 standard; protein; 1476 AA.
XX AC ABG32652;
XX DT 30-DEC-2002 (first entry)
XX DE P. luminescens (W-14) TcdB protein.
XX KW Gene; ds; toxin A; toxin B; TcdA; protoxin; TcdB; TccC2; transgenic;
XX KW monocot cell; dicot cell; oral toxin; insect; pest; TcdA.
XX OS Photothhabdus luminescens.
XX PN US2002078478-A1.
XX PD 20-JUN-2002.
XX PF 26-MAR-2001; 2001US-00817514.
XX PR 24-MAR-2000; 2000US-0191806P.
XX XX (FFRE/) FFRENCH-CONSTANT R H.
XX PA (BOWE/) BOWEN D.
XX PA (ROCH/) ROCHELEAU T A.
XX PA (WATE/) WATERFIELD N R.
XX FI Ffrench-Constant RH, Bowen D, Rocheleau TA, Waterfield NR;
XX DR WPI; 2002-655379/70.
XX DR N-PSDB; ABS52584.
XX PT Novel nucleic acid sequences which encode genes, tcdB and tccC2 from

PT Photothhabdus luminescens W-14, useful in heterologous expression of
XX orally active insect toxins.
PS Claim 1; Page 20-23; 40pp; English.
XX The invention discloses an isolated nucleic acid that encodes TcdB or
CC TccC2 from Photothhabdus luminescens W-14. Also disclosed is a transgenic
CC monocot or dicot cell and a transgenic plant (including the seeds) both
CC with genomes comprising tcdB and tccC2 nucleic acids. The nucleic acids
CC are useful for producing toxin A or B of P. luminescens W-14 in a
CC heterologous host and for encoding TcdB or TccC2 for producing an orally
CC active insect toxin in a host, where the host also expresses TcdA or TcbA
CC from P. luminescens W-14. Heterologous expression of Toxin A does not
CC afford the level of oral toxicity to insects as that of the native toxin,
CC but the coexpression increases this toxicity. The transgenic plants
CC expressing effective amounts of the toxins protect themselves from insect
CC pests. When the insects feeds on the transgenic plant it also ingests the
CC toxins and this deters the insect from further biting into the plant and
CC may even harm or kill the insect. The sequence presented is the P.
CC luminescens (W-14) TcdB protein
XX
SQ Sequence 1476 AA;
Query Match 76.3%; Score 6031.5; DB 5; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;
QY 1 MONSQDSFISITELSLPKGGGAI TGMGEAL TPTGPDGMAALS LPLPISAGRGYAPFTLNNY 60
DB 1 MONSOTFSVTLSL LPKGGGAI TGMGEAL TPA GPDGMAALS LPLPISAGRGYAPSLTLNNY 60
QY 61 SGAGNSPFLGWDGNCVMTIRRTTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114
DB 61 SGTGNSPFLGWDGNCVMAIRRTSTGPNVDETDTFLGPEGEVLVAALNEAGQADIRSES 120
QY 115 TLOGINLGATFTVTGYRSRLSESHFSRLSEYWKPTTKTDFTWLIYSPDQVHLLGKSPQAR 174
DB 121 SLOGINLGATFTVTGYRSRLSESHFNRLSEYWKPTTKTDFTWLIYSPDQVHLLGKSPQAR 180
QY 175 ISNPSQTTQTAQWLLEASVSRSRGEQIYYQVRAEDDTGCEADEITHHQAQRYLHIVY 234
DB 181 ISNPLNVNQTAAQWLLEASISHSRGEQIYYQVRAEDAGCETDELAHPSATVQRYLQTVHY 240
QY 235 GNRATSETLPGLDGSAPSOADWLFYLVDFDYGERSNNLKTTPAFSTTGSWLRCRDRFSRYE 294
DB 241 GNLTSADVFPFTLNGDDPLKSGMFCFLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFSRYE 300
QY 295 YGREIRTRRLCRQVLMYHHLQALDSKITENHGPTLVSRLLI LNYDESAITLAVFRRYGH 354
DB 301 YGFELRTRRLCRQILMFHRLQTLSGQAKGDDEPALVSRLLI LDYDENAMVSTLVSVRRYGH 360
QY 355 EQDGNVVTLPPELAYQDFSPRRHAWQPMQMDVLANFNAIQRWQLVLDLKGEGPLGLLYQDK 414
DB 361 EDNNTVTALPPELAYQDFPEPEQALWQSDVLANFNTIQRWQLLDLKGEGVPLGLYQDR 420
QY 415 GAWYRSARQLGEIGSDAVTWKMQPLSVTPSLQSNASLVNDINGDGLDWITGPELGRGY 474
DB 421 NGWYRSARQOAGEEMNAVWGMQLLPITPAVDQNASLMDINGDGLDWITGPELGRGY 480
QY 475 HSQRPDGSWTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAK 534
DB 481 HSQHPDGSWTRPTPLHALPIEYSHPRQAQLADLMGAGLSDLVLIGPKSVRLYVNRDRDGETE 540
QY 535 GKDVVQSGDITLPVPGADPRKLVAFSDVLGSGQAHVVEVSATKVTCPNLRGRFGQPIIT 594
DB 541 GRDVVQSGDITLPVPGADARKLVAFSDVLGSGQAHVVEVSATQVTCNENLGHGRFGQPIV 600
QY 595 LPGFSQPATEFNPQAQVYLADLDSGPTDLIYVHTNRDLDFLNKSGNGFAEPVTLRFPBGL 654
DB 601 LPGFSQSAASFPNDRVHLADLDSGPDALIVYHADRLDIFNKGNGFAKPTLSFPDGL 660
QY 655 RPDFTCQLQOMADVQGLGVASLIISVPHMSPHHWCDLTNMKFWLLNENNNNGVHHTLRY 714

Db 661 RPDDTQLOQVADYQGLGVVSLILSVPHMAPHHWRCDLTNAKPMWLLSETNNMGNATHLY 720
 Qy 715 RSSQFWLDEKAAALTTGTPVCYLPPPIHTLWQTEDEISGNKLVITLRYARGAWDGR 774
 Db 721 RSSVQFWLDEKAAALATGTPVCYLPPPVHTLWQTEDEISGNKLVITLRYARGAWDGR 780
 Qy 775 ERFGRFGVGYEQDSDHQAQGNAPERTPPALTKWVATGLPVDNALSTERYR-DQAFA 833
 Db 781 ERFGRFGVGYEQDSDHQAQGNAPERTPPALTATGKSWATGLPAVDNALSGYWGDKQAFA 840
 Qy 834 GFSPTFTWQNDKVDPLTPEDDNRYSWFNRALQGQLLRSELYGLDSDTNKHVYTYTEPR 893
 Db 841 GFTPTFTLWKGKVDPLTPEDDNRYSWFNRALQGQLLRSELYGLDGAQQIPIVTVESR 900
 Qy 894 SQVRLQHTDSRYPVLWSSVVERNHYERIASDPQCSNITLSSDRFGQPLKLSVQYP 953
 Db 901 PQVRLQDQATVSPVLWASVVERSYHYERIIISDPQCNQDITLSSDLFGQPLKLSVQYP 960
 Qy 954 RROQPAINLYPDTLPDKLLANSYDDQORQLRLTVQSSWHHLTNVTVRLGLPDRSRDI 1013
 Db 961 RRNKPTNPYDPLTPDLTPASSYDDQQLRLTCRQSSWHHLTGNEURLVLGLPDRSRDA 1020
 Qy 1014 FTYGAENVAGGLNELLSKNSLIADDPREYLQOQKATYTDQNTPTLQTPTRQALIA 1073
 Db 1021 FTYDAKQVPDGLNLETCAENSLIADDPREYLQOQKATYTDQNTPTLQTPTRQALIA 1080
 Qy 1074 FTETTVNQSTLAFNGISPSDKLSTLEAGYQOQNTYLFPRTGEDKVAHGYTDYGT 1133
 Db 1081 FTETAVLTESLLSAFDDGIIPTDELPGILTQAGYQOQEPYLPFRTGENKVMYARQGYTDYGT 1140
 Qy 1134 AAFQWRPKOSNTOLTKITLIDWANCYVVQTRDAAGLTTSYAKYDWRELTTPVQLTDIND 1193
 Db 1141 EAFQWRVPAQRNLSLLTGKMTLKWDTYCVITQTDAAGLTVSANYDWRFLTPQLTDIND 1200
 Qy 1194 NQHLITLDALGRPTTLRFWGTENGKMTGYSSPEKASFPSPSDVNAATLKKPLPVAQCV 1253
 Db 1201 NVHLITLDALGRPVQTRFMGIESGVATGYSSSEKPFPPNDIDTALNLTGPLPVAQCV 1260
 Qy 1254 YAPESNMPVLQKTFNRLAQDWKLYNARIITEDGRICHTLAYRWVQSOKAIPOLISLL 1313
 Db 1261 YAPDSNMPVLSQETFNLTQOEBETLDSRIITEDMWRICALTRRWLQSQKISTPLVKLL 1320
 Qy 1314 NNGPRLPPHSLTLTDRYDHDPEQIQQVVFSDGFCGLLQAAARHAGMARORNEPDSL 1373
 Db 1321 TNSIGLPPHNLTLTDRYDRDSEQIQQVAFSDGFCGLLQAAARHAGMARORNEPDSL 1380
 Qy 1374 IINVQHTENRMAVTGRTEYDNKQPIRTYQPYFLNDWRVYVNSDAROEKAYADTHYDP 1433
 Db 1381 VTKVENTKRWAVTGRTEYDNKQPIRTYQPYFLNDWRVYVNSDAROEKAYADTHYDP 1438
 Qy 1434 IGREIKVITAKGHRFTLTPFTVNEDENDTAAE 1468
 Db 1439 IGREIRVITAKGHRQSQFPFTVNEDENDTAAE 1473

RESULT 7

ADR21576
 ID ADR21576 standard; protein; 1476 AA.

XX AC ADR21576;
 XX DT 04-NOV-2004 (first entry)
 XX DE Photorhabdus TcdB1 toxin.
 XX KW toxin; insect; insecticidal; transgenic; pest control.
 XX OS Photorhabdus luminescens.
 XX PN W02004067727-A2.
 XX PD 12-AUG-2004.
 XX

PF 07-JAN-2004; 2004WO-US0000394.
 PR 21-JAN-2003; 2003US-0441723P.
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
 XX WPI; 2004-580999/56.
 DR N-PSDB; ADR21502.
 XX
 PT Controlling or inhibiting an insect, useful for pest control, comprises
 PT contacting the insect with effective amounts of a Protein A, a Protein B,
 PT and a Protein C.
 XX
 PS Claim 1; SEQ ID NO 22; 368pp; English.
 XX
 CC The invention relates to a novel method for controlling or inhibiting an
 CC insect comprising contacting the insect with effective amounts of a
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
 CC is encoded by a naturally occurring gene or has an amino acid sequence
 CC that differs from the product encoded by a naturally occurring gene only
 CC by truncation or by conservative amino acid changes. Protein A is a 230-
 CC 230 kDa toxin complex insect toxin that is derived from a first taxonomic
 CC species, has stand alone insecticidal activity, and has an amino acid
 CC sequence at least 40% identical to a sequence selected from XptAlwi,
 CC XptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin
 CC complex potentiator having an amino acid sequence at least 40% identical
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1xb,
 CC PotB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
 CC having an amino acid sequence at least 35% identical to a sequence
 CC selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, XptB1wi, XptC1xb, PptC1
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein
 CC C. The method is useful for pest control. The present sequence represents
 CC Photorhabdus luminescens TcdB1 toxin.
 XX

Sequence 1476 AA;

Query Match 76.3%; Score 6025.5; DB 8; Length 1476;
 Best Local Similarity 75.9%; Pred. No. 0;
 Matches 1119; Conservative 122; Mismatches 225; Indels 9; Gaps 3;
 Qy 1 MNSQSFITELSLPKGGGAIQMGALTPGDMALSLPLPISGRGYAPAFITLYN 60
 Db 1 MNSQSFITELSLPKGGGAIQMGALTPGDMALSLPLPISGRGYAPAFITLYN 60
 Qy 61 SGAGNSPFGMGWDCNVMTTTRRRTHFGVPHYDETDITFLGPEGEVLVVA-----DQPRDES 114
 Db 61 SGTGNSPFGMGWDCNVMTTTRRRTHFGVPHYDETDITFLGPEGEVLVVALEAGQADIRSES 120
 Qy 115 TLOGINLGATFTVTGYRSRLESFHSRLEYWQPKTKTGDFWLIYSPDQVHLIGKSPQAR 174
 Db 121 SLOGINLGATFTVTGYRSRLESFHSRLEYWQPKTKTGATDFWLIYSPDQVHLIGKSPQAR 180
 Qy 175 ISNPSQTTQTAQWILLEASVSSRGEQIYYQYRAEDDTGCEADEITHLQAQRYLHVY 234
 Db 181 ISNPLNVNQTAAQWILLEASVSSRGEQIYYQYRAEDDEAGCETDELAHAFSATVQRYLQVHY 240
 Qy 235 GNRATSETLPGDGSAPSQADWLFYLPVDFYGERSNLKTPTTAFSTTGSWLCRODRFSRYE 294
 Db 241 GNLITASDVFTLNGDDPLKSGWMLFCLVDFYGERNSLSEMPFKATGNWLCRDKDRFSRYE 300
 Qy 295 YGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLSRLLIYNDSEAIASITLVFVRVGH 354
 Db 301 YGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLSRLLIYNDSEAIASITLVFVRVGH 360
 Qy 355 EQDGNVVTLPPLLAYQDFSPRHAWQPMQDVLANFNAIQWOLVDLKGSLGLLYQDK 414
 Db 361 EDNNTVTLPLLAYQDFSPRHAWQPMQDVLANFNAIQWOLVDLKGSLGLLYQDK 420
 Qy 415 GAWWYRSQRGLBGTSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDVTITGFLRGY 474

DR	WPI; 1998-179427/16.	QY	651	PEGLRFDHTCOLQADVQGLGVASLILSVPHMSPHHRCOLTNMKPWLNNMNNNVH 710
XX	N-PSDB; AAV29924.	Db	660	PEGVQDNTCOLQVADIQSLGILSLITVPHAPHHRCOLSLTKPWLNNMNNNGAH 719
PT	Isolated toxins from Photorhabdus luminescens strains - useful for	QY	711	TLYRSSQFWLDEKAAALTTQOTPVCLYPPFPIHTLMQOTETDEISGNKLVTLLRYARGA 770
PT	control of insect pests.	Db	720	TLHYSSAQFWLDEKQLTKAGKSPACYLPFPFPHLLWYTEIQDEISGNRLTSEVNSHG 779
XX	Claim 34; Page 196-200; 321pp; English.	QY	771	WDGRRPFGVGVVEQTDHSHQLAQGNAPERTPPALTKNWYATCLPVIDNALSTEYWR-DD 829
CC	The present sequence represents a protein named TcaC of the bacterium	Db	780	WDGKERFPGFGICIKQDTTTFSHGTPAPQASPLSISWFGATGMDVDSQLATEYWOADT 839
CC	Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the	QY	830	QAPAGSPRFTTWO--DNKDVLTPBDDNSRYVFNRAKQGLRLSLYGLDDSTNKHVPY 887
CC	nematodes of the Heterorhabditis genus. The bacterium has at least 4	Db	840	QAYSGETRYTWDHTNQTDQAFTP-NETQRNWLTRALKQGLLRTLYGLDGTDKQTPY 898
CC	distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are	QY	888	TVTFPSQVRRLQHTDSRYVFLWSSVVEGRNHYHRIASDPQCSQNTISSDRFGAPLKQ 947
CC	produced from these regions that are associated with insecticidal	Db	899	TVSESYQVRSIPVNETELSAWVTAIENRSVHYERIITDPQFSQIKLQHDIFGQSLSQ 958
CC	activity. The native toxins are secreted proteins. The proteins are toxic	QY	948	LSVQYPRRQOPALNLYPDTLPDKLLANSYDDQORQLRLTYQSSMHLTNTNVTYVGLPD 1007
CC	to insects upon exposure and especially when ingested. The nucleic acid	Db	959	VDIAMPREKPAVNPYPPTLPETLPDSSYDDQQLRLVRQKNMHLTDTGENWRLGLPN 1018
CC	sequence can be used to produce transgenic plants, baculoviruses or	QY	1008	STRSDIFTYCAENVNYPAGLNLELLSDKNSLIADDKPREYLGQOKTAYTDGQNTTLPQTPT 1067
CC	microbial hosts for toxin production. They can be used to control insects	Db	1019	AQRDVTYVDRSKIPTEGILSLLEILLKDDGLLADEKAAYVILGQOQTFYTAGAEVLEKPT 1078
CC	pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,	QY	1068	RQALIAFTTETVNTQSTLSAFNGSIPIKSLSTLEQAGYQNTNLYPRTGEDKVVAAHGG 1127
CC	Dictyoptera, Acarina or Homoptera orders, especially the Southern or	Db	1079	LQALVAFQETAMMDTSLQAYEGVIEBEQELNTALTQAGYQQVAVLFNTRSESPVMAARQ 1138
CC	Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,	QY	1128	YTDYGTAAQFWRQKQNTQLTKITLIWDANYCVVVTOTDAAGLTTSKYDWRRLTPVQ 1187
CC	turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,	Db	1139	YTDYDAAQFWRQKQNTQLTKITLIWDANYCVVVTOTDAAGLTTSKYDWRRLTPVQ 1198
CC	corn earworm, European corn borer or tobacco hornworm or budworm	QY	1188	LTDINDNQHLITLALGRPITLRFMGTEGKMTGYSSPKASPSPPSDVNAALTELKPLP 1247
XX	Sequence 1485 AA;	Db	1199	LTDINDNQHLITLALGRVTTSEFGTEAGAAAGYSN--QPTPDPDVKALALTGALP 1255
Query Match	58.0%; Score 4581.5; DB 2; Length 1485;	QY	1248	VACQVYAPESMMPVLSQKTFNRL---ABQDWOKLYNARIITEDGRICITLAYRWVQSOK 1304
Best Local Similarity	57.6%; Pred. No. 0;	Db	1256	VACQVYAVDSWMPVLSQKTFNRL---ABQDWOKLYNARIITEDGRICITLAYRWVQSOK 1315
Matches	858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;	QY	1305	AIPQLISLNNPRLPHSLITLTDYDHDPEQOIRQVVPSPDGRGLLQAAARHAGMA 1364
1	MONSDPSITSLPCKGGGATGGEALTPGDMGMAALSPLPISAGRYAPAFNTLYN 60	Db	1316	LTIQLISLNNPRLPHSLITLTDYDHDPEQOIRQVVPSPDGRGLLQAAARHAGMA 1375
1	MQDSPSVISITSLPCKGGGATGGEALTPGDMGMAALSPLPISAGRYAPAFNTLYN 60	QY	1365	RQNEEDGSLIINVQ-----HTENRWAVTGRTEYDNKGOPRTYQYFVFNDRWYVNSDA 1418
61	SGAGNSPFLGWCNVMTIRRTTHFGVPHYDETDFTFLGPEGEVLVVA-----DQPRDE 113	Db	1376	WQKEDGGLVVDANGVLVSAPTDRNAVSGTEYDDKGPVRYQYFVFNDRWYVNSDA 1435
61	NSAGNGPFGIGWCNVMSISRRTHQGIPOYQNDGDTFLSPQGEVNMIALNDQGPDIRQDV 120	QY	1419	RQKEAYADHTVVDTPGREIKVITAKGWFRFTLTFTWFTVNEDENDTAA 1467
114	STLQINGLNGATFTVGRSLESFHSFSLVWQPKT--TGKTDWFLIYSPDGQVHLGKSP 171	Db	1436	RD--DLFADTHLYDPLGREYKVITAKKLYREKLYTPFWFVSEDENDTAS 1482
121	KTLQGVTLPISYVTRYQARQLDLSKIEYVQWSPASQGEGR--FWLISDPDHLILKTA 179	RESULT 11		
172	QARISNPSTTTQAQWLLASVSRGQIYYQYRAEDDTGCEADBEITHLQATAQRYLHI 231	AAW17887		
180	QACLANPQNDQIAQWLLSTVTPAGEHVSQYRAEDEAHCDNNEKTAHFNVTVAQRYLVQ 239	ID	AAW17887	standard; protein; 1485 AA.
232	VYGNRTASTLPGLDGSASQADWLFYVDFVGRSNNLKTTPAPST--TGSWLCQDRF 290	XX	AAW17887;	
240	VNYGNIKPQASFLVDNAPPAPEEWLFPLVDFHGERDTSLHTVPTWDAGTAQMSVPRDIF 299	XX	AC	
291	SRVEYGEPIRTRRLCROVLMYHLQALDSKITEHNGPTLVSRLLIYDESAIASTLVFVR 350	XX	AC	
300	SRVEYGEVTRRLCQVLMFHRALMAGASTNDAPELVGRLLILEYDKNASVTLLITIR 359	DT	17-OCT-2003	(revised)
351	RVGHEQDGNVTVLPPLLAYQDPSRPHAHQWPMVLANFNALQWQLVLDKGEGLPLG 410	DT	29-JAN-1998	(first entry)
360	QLSHESDGRPVTPQPLSLAWQRPDLKIPVQWRFADLNDNFNSQOQYQLVLRGEGPLM 419	XX	DE	
411	YQDKGAWYSRQBLGIGSDAVTWKQPLSVIPSLQSNASLVVDINGDQOLDWITGPG 470	XX	XX	Photorhabdus luminescens insect toxin protein TcaC.
420	YQDRGAWYKAPQQRQDGSNAVYDYKIAFLPTLPNLQDNASLMDINGDQOLDWVVTASG 479	KW	KW	Insecticide; insect; toxin; pest control; biological control;
471	LRGHVSORPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530	KW	KW	Photorhabdus luminescens; TcaC; Southern corn rootworm;
480	IRGHVSQOPGKWTFTFIPALPVEYFHPISQFADLTGAGLSDLVLIGPKSVRLYANQRN 539	KW	KW	Colorado potato beetle; Western corn rootworm; meal worm; boll weevil;
531	GFANGKDVQSGDITLVPVGPADPKLVAFSDVLGSGQAHLVEVSATKVTCPNIGRGRPG 590	KW	KW	turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper;
540	GWRKGEDVPQSTGITLPTVTGTDARKLVAFSDMLGSGQQHLVEIKGNRVTCPNIGHGRFG 599	KW	KW	codling moth; corn earworm; European corn borer; tobacco hornworm;
591	QPTLPFGSQPATEFNPQAVYLADLDCSGPTDLYVHTNRDLDFLNKSGNGFAPVTLRF 650			
600	QPLTSLGFSQPENSFNPERFLADIDSGTGTDLIYAQSGSLIYLNQSGNQFDAPLTLAL 659			

KW tobacco budworm; Lepidoptera; Hymenoptera; Diptera, Dictyoptera; Acarina; Homoptera.

XX OS Photorhabdus luminescens; strain W-14 (ATCC 55397).

XX PN WO9717432-A1.

XX PD 15-MAY-1997.

XX PF 06-NOV-1996; 96WO-US018003.

XX PR 06-NOV-1995; 95US-0007255P.

XX PR 28-FEB-1996; 96US-00508423.

XX PR 28-AUG-1996; 96US-00705484.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX PA Ensign JC, Bowen DJ, Petell J, Patig R, Schoonover S;

XX PI Ffrench-Constant RH, Rocheleau TA, Blackburn MB, Hey TD, Merlo DJ;

XX PI Orr GL, Roberts JL, Strickland JA, Guo L, Ciche T;

XX WPI; 1997-281022/25.

XX N-PSDB; AAT68840.

XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them - can be genetically engineered into insect larvae food and plants for insect control.

XX Claim 34; Page 159-164; 276pp; English.

XX This polypeptide comprises a specifically claimed insecticidal toxin protein, TcaA, of Photorhabdus luminescens, a 184 kDa a component of a toxin protein complex. TcaB can be expressed in host cells using a gene (see AAT68837) isolated from a genomic library. Claimed toxin proteins of P. luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or genetically engineered into, insect larvae food and plants for insect control. The Photorhabdus toxins are particularly effective against southern corn rootworm, Colorado potato beetle, Western corn rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling moth, corn earworm, European corn borer, tobacco hornworm and tobacco budworm (Lepidoptera), and are also active against insects of the orders Hymenoptera, Diptera and Dictyoptera, Acarina and Homoptera. (All claimed). (Updated on 17-OCT-2003 to standardize OS field)

XX SQ Sequence 1485 AA;

Query Match 58.0%; Score 4580.5; DB 2; Length 1485;

Best Local Similarity 57.6%; Fred. No. 0;

Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

Qy 1 MONGDPSITSLPKGGGALTGMEALTPPTGPMMAALSPLPISAGRGVAPAPFTLNYN 60

Db 1 MQDSPEVSITSLPKGGGALNGEALNAAGPDMASLSPLPSTGRGTAPGLSLIYS 60

Qy 61 SGAGNSPFLGWCDCNVMTIRRTFHGPHYDETDTFLGPEGEVLVVA-----DQPRDE 113

Db 61 NSAGNGPFGICWQGVMSISRTQHGIPQYGNDDTFLSPQCEVMNIALNDQGPDIRQDV 120

Qy 114 STLQGINLGATFTVTGRSLRLESFHSRLLEYKQPKT--TGKTDFTMLIYSPDQVHLLGKSP 171

Db 121 KTLQGVLPISYTVTRYQARQILDPSKIEYVQPASGQEGRA-FWLSTPDPGLHLGKTA 179

Qy 172 QARTSNPQTQTQAWLLEASVSRGEQIYYQYRAEDDTGCEADBITHLQATQRYLHI 231

Db 180 QACLANPQNDQIAQLLEETVTPAGEHVSQYRAEDBAHCDNEKTAHFNLTQRYLVQ 239

Qy 232 VYGNRTASETLPLGDSAPSQADWLFYLVFDYGERSNLNKTPPAFST-TGSMCLCRDPRF 290

Db 240 VNYGNIKPQASLFLVLDNAPPAPPEWLPHLVFDHGERDTSHTVPTWDAGTAQMSVRPDI 299

Qy 291 SRYEYGFIRTRRLCRQVLYMTHHQLQALDSKITEHNGPTLVSRLLINLYDESAIASTLVFR 350

300 SRYEYGFVTRRLCQOVLMFHRTALTALMAGEASTNDAPELVGRLLILEYDKNASVTTLITIR 359

Qy 351 RVGHEQDGNVVTLPPLLELAYQDFSPRHHAHQWMDVLANFNAIQRWQLVLDKSGELGCLL 410

Db 360 QLSHESDGRPVTPPPLLELAQWRFDELEKIPWQRFDALDNFNSQQRYYQLVLRGSGLPGL 419

Qy 411 YQDKGAWYRSARQLBEGISDAVTWEKQKPLSVIPSLQSNASLVLDINGDQOLDWITGPG 470

Db 420 YQDRGAWYKAPQRBQEDGSDNAVYTKIAFLPTLPNLQDNASLMDINGDQOLDWVVTASG 479

Qy 471 LRGYHSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLDLVLGPKSVRLYANTRD 530

Db 480 IRGYHSQOPDGKWTFTFIPNALPVEYTHPRAQLADLMGAGLSLDLVLGPKSVRLYANTRD 539

Qy 531 GFAKGDVWQSGDITLPPVPGADPRKLVAFSDVLGSGQAHLVESATKVTCPNGLGRGFRG 590

Db 540 GWRKGEDVPQSTGITLPVTGTDAKLVAFSDMLGSGQOHLVEIKGNRVTCWPNLGHGRFG 599

Qy 591 QPITLPGFSPATFEPNPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRF 650

Db 600 QPLTSGFSQFENSFNERFLADIDGSGTDTLIYAQSGSLIYLQNSGQFADPLTLAL 659

Qy 651 PEGLRPDHTCOLQWADVQGLGVASLILSVPHMSPHWRCDLTNNKPLLNNMNNMVGHH 710

Db 660 PEGVQFDNTCOLQVADIQGLIASLITVPHIAPHWRCDLSTKPKLLNVMNNRGAHH 719

Qy 711 TLYRSSSQFWLDEKAAALTTGQTPVCYLPFPPIHTLMQTEDEISGNKLVITLRYARGA 770

Db 720 TLYRSSAQFWLDEKQLTKRAGSPACYLPFPWLLWYTRIQDEISGNRLTSEVNYSHGV 779

Qy 771 WDGRERFRGFGVVEQDTHQLAGNAPPTPALTKNWTATGLPVIDNALSTELTYR-DD 829

Db 780 WDGRERFRGFGCITKQDITNFSHTAPEQAAPSLISWFAFGMDEVDQSOLATEYQWADT 839

Qy 830 QAAGSPRFTTQ--DNKDVPLTPEDNRSYVFNFRALKGQLLSSELYGLDDSTNKVPEY 887

Db 840 QAAGSPRFTTQ--DNKDVPLTPEDNRSYVFNFRALKGQLLSSELYGLDDSTNKVPEY 887

Qy 888 TVTEFRSQVRLQHTDSRYVPLWSSVVESSNYHYRIASDPQCSQNTLTSDFRGQLKQ 947

Db 899 TVSESRVQVRSIPVKNKTELAWTAIENRSYHYRIITDPQRSQSIKLQHDIFGQSLOS 958

Qy 948 LSVQYPRQOPAINLYPDTLPDKLLANSYDDQRLALTYQQSSWHLLTNTVTVLGLPD 1007

Db 959 VDIAMPRERKPAVNPYPPTLPETLFDSSYDDQQLLRLVRQKNSHLLTDGENWRLGLPN 1018

Qy 1008 STSDIFTYCAENVPAAGLNLLELSDKNLSIADDPREYLGQOKTATYDQNTTPTQET 1067

Db 1019 AQRDRVTYDRSKIPTTGISLEILLKDDGLLADKAAVILGQQQTFYTAGQAEVTLKPT 1078

Qy 1068 RQALIAFTETTVFNQSTLSAFNGSIPSDKLSLTLLEQAGYQQTNYLPFRTGDKVWVAHG 1127

Db 1079 LQALVAFQETAMDDNSLQAYEGVIEQELNTELTQAGYQOVARLFWTRSESFWAARQ 1138

Qy 1128 YTDYGTAAQFWRPQKQNTOLTGKILTIWIDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQ 1187

Db 1139 YTDYGTAAQFWRPQKQNTOLTGKILTIWIDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQ 1187

Qy 1188 LTIINDNHLITLDALGRPITLFWGTENGKMTGYSPEKASFPSPDVAALTELKPLP 1247

Db 1199 LTIINDNHLITLDALGRVTTFRFWGTENGKMTGYSPEKASFPSPDVAALTELKPLP 1255

Qy 1248 VAOCQVYAPESWMPVLSQKTFNRL---AEQWOKLYNARIITEDGRITCLAYRWRVQSOK 1304

Db 1256 VAOCQVYAPESWMPVLSQKTFNRL---AEQWOKLYNARIITEDGRITCLAYRWRVQSOK 1304

Qy 1305 AIPQLISLANNGRLPHSLTLTTRDYDHPDEQOIRQOVVFSFGFGLLQAAARHEGMA 1364

Db 1316 LTIQLISLLASIPRLPHVLIITDRYDSDPQQOQHTQVFSFGFGLLQAAARHEGMA 1375

Qy 1365 RORNEDESLINVO-----HTENRMAVTGRTEDNKGOPITRYQPYFLNDRWYVSDSA 1418

Db 1376 WQRKEDGGLVVDANGVLVLSAPTDTTRWAVSGRTEYDDKGPVRYQPYFLNDRWYVSDSA 1435

Db 1133 PQSDVAGSSKVVWVARQYTEYGSAQFYRPLIQRKSLLTGKVTLSWDTHYCVVVKTED 1192

Qy 1169 AAGTTTSKDYRFLTPVQLTDINDNQHLLTDLALGRPITLRFWGTENGKMTGYSSPEKA 1228

Db 1193 GAGMTTQAKYRFLPQALTDINDNQHIVTFNALGQVTSRFGWTENGKISGYSTPEK 1252

Qy 1229 SFSRPSDVNAIELKKPLPVACQVYAPESMPVLSQTFN---RLAEQDQKLYNARI 1285

Db 1253 PFTVPDTEKALALQFTIPVSCQNIYVDSWMLLPQOQLTGQKKEGTELLNALLHRAGW 1312

Qy 1286 TEDGRICITLAVRRVWQSKAIPQ---LISLNNGPRLPPHSLTLTDDRYDHDPEQOIRQ 1341

Db 1313 TEDGLICELAVRMIKQATSMMAVTLQOILAQTPROPFHAWITTDYDSDSQOQLRQ 1372

Qy 1342 QVPSDGFGRLLQAAARHAGWARQORNEGDSLIIN-----VQHTENRWAVTGRTEYDNK 1395

Db 1373 SIVLSDGFGRVLQSAQRHEAGEAWQRAEDGSLVVDNTGKPVVANTTTTAWASVGRTEYDGK 1432

Qy 1396 GQPIRTQPYPLNDWRVYNSDSARQEKAYADTHVYDPIGREIKVITAKGFRRLTFPW 1455

Db 1433 GQAIRAYLPYPLNDWRVYNSDSDARD--DLAYADTHFYDPLGREYQVKTAKGFWRNMFMPW 1490

Qy 1456 FTVNEDENDTAAEV 1469

Db 1491 FTVNEDENDTAAEL 1504

RESULT 13

ABB47812

ID ABB47812 standard; protein; 1506 AA.

XX ABB47812;

AC ABB47812;

XX 22-SEP-2005 (first entry)

DE Native XptB1(xb).

XX XptB1; toxin; insect resistance; insecticide.

XX Xenorhabdus bovienii.

XX US2005155104-A1.

XX 14-JUL-2005.

XX 23-DEC-2004; 2004US-00020848.

XX 07-JAN-2004; 2004US-0534893P.

XX (APEL/) APEL-BIRKHOED P C.

PA (HEYT/) HEY T D.

PA (THOM/) THOMPSON R L.

PA (MEAD/) MEADE T.

PA (LIZS/) LI Z S.

PA (RUSS/) RUSSELL S M.

PA (SHEE/) SHEETS J J.

PA (LIRA/) LIRA J M.

PA (FENC/) FENCIL K J.

PA (MITC/) MITCHELL J C.

XX

PI Apel-Birkhold PC, Hey TD, Thompson RL, Meade T, Li ZS;

PI Russell SM, Sheets JJ, Lira JM, Fencil KJ, Mitchell JC;

XX WPI; 2005-496874/50.

DR N-PSDB; ABB47811, ABB47817, ABB47819.

XX

DR New isolated protein and encoding nucleic acid having toxin activity

PT against an insect, useful for developing new insecticidal toxins that can

PT be used to control insects.

XX

PS Claim 3; SEQ ID NO 2; 54pp; English.

XX

CC This sequence represents native XptB1(xb). XptB1 is one of five TC (toxin

CC complex) proteins from Xenorhabdus. XptA1 is a "stand alone" toxin. XptA2

CC also has some stand alone toxin activity. XptB1 and XptC1 are the

CC xenorhabdus potentators that can enhance the activity of either

CC both of the XptA toxins. XptD1 has some level of homology with TcCB

CC (toxin complex c protein B). This protein has toxin activity against an

CC insect. The polynucleotide encoding this protein may be used to transform

CC a plant cell, and thereby generate a transgenic plant which may be

CC ingested by an insect, whose numbers are subsequently controlled. The

CC methods and compositions of the present invention are useful for

CC developing new insecticidal toxins and other proteins that can be used to

CC control insects.

XX

SQ Sequence 1506 AA;

Query Match 53.9%; Score 4261; DB 9; Length 1506;

Best Local Similarity 54.1%; Pred. No. 0;

Matches 819; Conservative 228; Mismatches 409; Indels 58; Gaps 21;

Qy 2 QNSQDFSIHELSPKGGGAIITGCEALTPTGPGMAALSPLPLISAGRGVAPATLYNVS 61

Db 3 QDSQDMVTVTQLSLPKGGGAIISGMGDTISNAGPDGMAASLVPLPISAGRGGAPNLSLYSS 62

Qy 62 GAGNSPFGLGWDCNVMTIRRTTHFGVPHYDTEDTFLGPEGEVLVA---DQP--RDEST 115

Db 63 GAGNSGFGIGWQSGSTMAISRTQHGVEQYHGEDTFLCPMGEVMAVANQSQQDPVRKTDK 122

Qy 116 LQGINLGATFTVTGYRSRLESFHSRLEYWQPKT--TGKTDFWLIYSPDQVHLLGKSPQAR 174

Db 123 LGGQLPVTVTTRHQPRNIQHFSKLEYWQPTDVTTPFWLMYSPDQIHIFGKTEQAQ 182

Qy 175 ISNPSQTTQTAQMLLEASVSRGEOIYYQRAEDDTGCEADEITHHQAQRYLHVYY 234

Db 183 IANPAEVSQIAQWLLBEETVTPAGEHIYYQYRAEDDTCDDSEKNAHPNASEAQRILTQVNY 242

Qy 235 GNRATSETLPGDGSAPSQADWLFYLDYGERNNLKTTPAR--STGSMLCRODRSRV 293

Db 243 GNITPESLLVLKNTPPADNEWLFHLVDYGERAQEINTVPPFKAPNNMKIRDFRSRF 302

Qy 294 EYGFETRRLCRQVLMYHHLQAL--DSKITEHNGPTLVSRLLILNYDESASIASTLVFVRV 352

Db 303 EYGFETRRLCQQLLMFHLKSLAGEIDGEEIPALVANLLSYDLNDSVTTLTAIRQM 362

Qy 353 GHEODGNVTLPLLELAYQDFSPRHHAWQPMVLANFNALQRMQVLDLKEGELPGLLYQ 412

Db 363 AYETDATALIAPLEFDYQPFPEAKVTQKWOEMPQAGLNAQQPYQLVDLYGEGISILYQ 422

Qy 413 DK-GAWVRSQAQRLGEIGS--DAVTWEKMQPLSVTPSLQSNASLVINDGCGOLDWVITGPG 470

Db 423 DRPGAWWYQAPIROKQNVEDINAVTYSFINPLPKPQQDDRATLMDIDGDLHDMVIAGAG 482

Qy 471 LRGYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530

Db 483 IQGRYSMQPNGEWTHFIPISALPTEYFHPQALADLVGAGLSDLALIGPSRVLYANDRG 542

Qy 531 GFAGKDVQSGDITLVPQADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNPLGRGRFG 590

Db 543 NWRAGINVMPPDGVNLEIFGCDASSLVAFSDMLSGGQQLHVEIAAQSVCWPNLGHGRFG 602

Qy 591 QPITLPGFSQPAETEFNPAQVYVLADLDSGPTDLYVHTNRDLIFLNKSGNGFAEPVTLRF 650

Db 603 AAILPGFSQPNGTGFNANQVFLADIDGSGTADIIYAHSTYLDIYLNESGNRFSAPVRLNL 662

Qy 651 PEGLRFDHTQLOMADYVQGLGVASLILSVPHMSPHWRCDLTNNKPLNEMNNMNVH 710

Db 663 PEGVWFNTQLOVSDIQGLGAASIVLTVPHMTPRHWRYDFTHNKPWLLNINNRRGNET 722

Qy 711 TLYRSSQFWLDEKAAALTTGQPPVCYLPPPIHTLMQTEDETSIGNKLVTLTLRYARGA 770

Db 723 TLFYRSSAQFWLDEKSKQIEELGKFAASYLPPLPPIHLLWRNEALDEITGNRLTKVNVYAHGA 782

Qy 771 WDGEREPRGEGYVEQDTHQLAQGNAPERTP-----PALTKWNYATGLPVIDNALSTEY 825

Db 783 WDGEREPRGEGYVEQDTHQLAQGNAPERTP-----PALTKWNYATGLPVIDNALSTEY 841

Qy	826	WR-DDQA	PAGFSRFRFTW-----QDNKVPLPDPEDDNSRYWFNRALKGQLLRSELYGLD	878
Db	842	WRGDOA	FAGTFRFTRYEKGNAGCEQODTFPIKEPTETRAYWLNRAMKQGQLLRSEVYG-D	900
Qy	879	DSFNK-HVPY	TTFPSROVERLQHTOSRYPVYLWSSVVSESRYHYERTASPOCSQNITLS	937
Db	901	DKTEKA	KIPVTVEARCOVRLLIPSNDAAAPSSWTSSIENRSYHERIVDPSPCKQQVVLK	960
Qy	938	SDRFOP	LKQLSVOYPRRCOOPAINLYPDTLFDKLLANSYDDQORQLRLTYQQSSWHHLTN	997
Db	961	ADEYGF	PLAKVDIAYPRENKPAQNPYPDSLPTLFADSYDDQQQLVLTQQOQSYHLLTQ	1020
Qy	998	NTRVRL	GPLPSTSRSDIPTY----GAENVPAGLNLELLSDKNLSLIADDKPREYLGQOKTA	1053
Db	1021	QDDWVL	GLTDSRYSEVHYAQTDQAQSDIPKAGILTEDLLXKVGDLIGKDXFTFVILGOORVA	1080
Qy	1054	YTQONT	PTLPQTBRQALIAFTTTVPNQSTLSAFNGSIIPSDKLSTLEOGAYOOTNYLF	1113
Db	1081	YVGGD----	AEKPTRQVRVAYTETAAPDDNHALHAFDGVIADELTOQLLAGGY-----LLV	1132
Qy	1114	PRTGE-----	DKVMVAHHGYDTYGTAAQFWRPQOKSNTOQLTGKITLLWDANYCVVUOTRD	1168
Db	1133	PQISDV	AGSSEKVMVARQGYTEYGSAQFYRPLIQRKSLITGXKYLTSMDTHYCVVKVED	1192
Qy	1169	AAGLTTS	AKDWMLPTPVOLTDINDNOHLITLDALGRPYLTRFWGTENGKMTGVSSPEKA	1228
Db	1193	GAGMTT	OAKDYDRFLPQAULTDINDNOHVTFNALGQVTSRFGWTEGKISGYSTPESK	1252
Qy	1229	SFPSPS	DVNAAIELKPLPVAQCQVYAPESMWPVLISKTFN---RLAEQDWQOKLYNARI	1285
Db	1253	PFTVPD	TEKALALOPIPVSOQNIYVPDSWMRLLPQOULGTQKGETIWNALHRAGVV	1312
Qy	1286	TEDGRIC	TLAYRRWVSQKAIPQ----LISLLNNGPRLPPhSLTLTTDDYHDHPHQOIRQ	1341
Db	1313	TEDGLI	CELAYRWIKRQATSSMAVTLQILAQTPROPPHAMTITTTDRYDSDSQOQLRQ	1372
Qy	1342	QVVPDG	PGRLLOAAARHEAGMARONEDSLLIN-----VQHTENRWAVTGETYDNK	1395
Db	1373	SVLSDF	GVRLVQSQRHEAGWARABDGSLVVDNTGKPVAVNTTTFRWAVSGRTEYDGK	1432
Qy	1396	GQPIRT	TPYFLNDWRVYSNDSARQEKEAVADTHVDPIGREIKVITAKGWFRRTLTPPW	1455
Db	1433	GQAIRA	LYPYLNDWRVYSDSARD--DLLYADTHFYDPLGREYQVKTAKGFWRNMFMFW	1490
Qy	1456	FTVNED	ENDATAEV	1469
Db	1491	FVYNED	ENDATAARL	1504

RESULT 14

AA097695	AA097695 standard; protein; 1428 AA.
XX	
XX	AA097695;
XX	
XX	
DT	19-JUN-2001 (first entry)
XX	
DE	SePB protein encoded by <i>Serratia</i> insecticidal protein complex gene.
XX	
KW	Insecticidal protein complex; amber disease; insect; Coleoptera;
KW	pesticide; SePB protein.
XX	
OS	<i>Serratia</i> sp.
XX	
PN	WC0200116305-A2.

PA	(NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.	
XX		
PI	Glare TR, Hurst MRH, Jackson TA;	
XX		
DR	WPI; 2001-169009/17.	
DR	N-PSDB; AAA91292.	
XX		
PT	New nucleic acid encoding a polypeptide useful as a pesticide especially	
PT	for Coleoptera.	
XX		
PS	Claim 24; Page 100-105; 109pp; English.	
XX		
CC	This sequence represents the SepB protein encoded by the Serratia	
CC	insecticidal protein complex gene of the invention. The invention relates	
CC	to a gene encoding an insecticidal protein complex or a functional	
CC	fragment, a neutral mutation, or a homologue of the complex. The	
CC	polypeptides and nucleotides of the invention are used to induce amber	
CC	diseases or like conditions in insects, especially to those from the	
CC	order comprising Coleoptera, useful as a pesticide	
XX		
SQ	Sequence 1428 AA;	
	Query Match	50.3%; Score 3971; DB 4; Length 1428;
	Best Local Similarity	52.7%; Pred. No. 2.3e+309;
	Matches	783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;
QY	1 MONSQPSITELSLPKGGGAITCGEALPTGTGPDGMAALSPLPIAGAGYAPAFTLN	60
DB	1 MQNHQDAITAPTLPGGGAVTGLKGDIAAGPDGAATLSIPLVSPGKGAYATGALN	60
QY	61 SGAGNSPFGLGWDCNVTMTRRRTHFGVPHYDEDTFLGPEGEVLVA-----DQ	114
DB	61 SRSGNGPFGIGWGIGAAVQRRTRNGAPTYYDDTDEFTGPDGEVLVPAALTAAG	120
QY	115 TLGINIGATFTVTGTRSLRSHFSRLWYQPKTKTDFWLYSPDQGVHLLGKSPQAR	174
DB	121 SLAIGIPGGSFNQVYRSRTEGSLRSLRWLPADETETEFVLYTDPGQVALLGRNAQ	180
QY	175 ISNPSQTTQAOALLASVSSRCGLYYQVRAEDDTGCCADEITHLQATAQRYLHVY	234
DB	181 ISNPTAPTQAVMLMESSVSLTGEQMYQYQRAEDDDGCEAEEDAHAPQAQRYFVA	240
QY	235 GNRTASETPLGLDGSAPSQADWLFYLVFDYGERSNNLKTTPAFSTTGS--WLC	292
DB	241 GNRQAARTLPAL-VSIFSDMSWLFILVFDYGERSSVLSEAPAWQTGSGWMLCRQ	299
QY	293 YEGFEIRTRLRCQVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESAIATSLV	352
DB	300 YEGFNLIRTRLRCQVLMFHYLGLAGSSGANDAPALISRLLLDYRESPSLSLLEN	359
QY	353 GHQDQGNVVTLPLELAYQDPSRHHAHQPMVDVLANFNAIQRWQLVDLKGELGLLYQ	412
DB	360 AYSDGTSCLPALALAGWQTFPTPTLSAWQTRDDMGKLSLLQYQLVDLNGE	419
QY	413 DKGAWYRSQAQLGEIGSDAVTWEKMQPLSVIPSLOSASLVINDINGDGLDWTIT	472
DB	420 DSGAWYREPVROSGDDPDVATWGAAALPTMPALHNSGLIADNGDLREWVWTA	479
QY	473 GYHSQRPDGSWTRFTPLNALPVSYTHPRAQLADLMAGLSDLVLI GPKSVRLYAN	532
DB	480 GWYDRTPGRDNLHFTPLSALPVEYAHPKAVLADILGAGLTDMVLIGPRLVSY	539
QY	533 AKGDVQVSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVVSATKVTCTWNLGR	592
DB	540 NKGETVQQTRELTLPVPGVDPRTLVAFSDMAGSGQQLTEVRANGVRYWPNL	599
QY	593 ITLPGFSQATEFNPAQVYLADLDGSGPTDLIVVHTNRLDIFLNKSGNGFAEP	652
DB	600 VNIPGFSQSVTTNPDQILLADTDGSGTTDLIYAMSDRLVIYFNQSGNYFAE	659
QY	653 GLRPDHTCQLQMAADVQSLGVASILSVPHMSPHHWRCDLTMMKFWLLNEMNN	712
DB	660 GVRDYRTCSLOVADIOSLGVPSSLLTVPVAPHHVCHLSADKFWLLNGNNN	719

```
Qy 713 YRSSQFWLDEKAAALTTGTPVCYLPPFIHTLMQTEDEISGNKLVTLRYARGAWD 772
Dy 720 HYRSSVQFWLDEKAAALAGSPACYPFLTLHLWRSVVDITGNRLVSDVLYRHGWMD 779
Qy 773 GRERFRGFGVVEQTDHOLA-CGNAPERTPALTKNMYATGLPVDINALSTEWYRDD-Q 830
Dy 780 GQERFRGFGVVEQTDHOLA-CGNAPERTPALTKNMYATGLPVDINALSTEWYRDD-Q 839
Qy 831 AFAGFSRFTTQDNKDVPLTFPEDNSRYFNWNRKALGQLLRSSELYGLDDSTNKHVPYTVT 890
Dy 840 AFADFAFTFTVSGEDQYTP-DDSKTFWLQRLKGLLESELYGADGSSQADIPISVT 898
Qy 891 EFRSQVRLQHTDSRYVPLWSSVVESSNYHYERIASDPQCSQNTLSSDRPGQPLQSLV 950
Dy 899 ESRPQV-RLVEANGDYPVWPMGAESRTSVYRYHNDPQCCQQAVALLSDEYGFPLRQVSV 957
Qy 951 QYPRRQPAINLYPDTLPDKLLANSYDDQQLRLTYQSSWHHTNTNVR--VLGLPDS 1008
Dy 958 NYPRRPSADNPYPASLPATLAFANSYDEQQOILRLGLQSSAHLVLSLSEGHVLLGLAEA 1017
Qy 1009 TRSDIFYGAEVNPAGGNIPELLSDKNSLIADDPREVLGQOKTAYTDGQNTTLPQTPT 1068
Dy 1018 SRDVFYSDADNPVEGGLLEHLLAPESLVSQVGLTAGQQVWYLDSDQVATVAAPPL 1077
Qy 1069 QALIAFTTTFVNSTLSAFNGSIPSKLSTLLEQAGYQNTNLYFP--RTGEDKVVVAHH 1126
Dy 1078 PPKVAFIETAVLDGEMVSSLAAYIVDEH---LEQAGYRQSGYLFPRGREAEQALWTQCQ 1133
Qy 1127 GYTDYGTAAQFWRPQKQSNQTLTKITLINDANYCVVQTPDAAGLTSAKYDWRFLTPV 1186
Dy 1134 GYVYAGAEHFWLPLSRDMSMLTGPVITRDAYDCVITQWQDAAGIVTADYDWRFLTPV 1193
Qy 1187 QLTIDINQHLITDALGRPITLFWGTENGKMTGYSSPEKASPPSDVNAALTELKPL 1246
Dy 1194 RVTDPNDNLQSVLLDGLRVTLFWGTENGIAATGYS--ATLSVPDGAALALATPL 1250
Qy 1247 PVAQCQVAPESMMPVLQSKTFNRLABQDWOKLYNARIITEDGRICTLAYRRVQSQAI 1306
Dy 1251 PVAQCLVVTDSW-----GDDNEK----- 1270
Qy 1307 POLLSLNNGRPLPHSLTLTDRYDHPDQOIRQOVVFSDFGRLLOAAARHAGMARQ 1366
Dy 1271 -----MPPHVVLATDRYSDTGQOVRQOVTFSDGFGRELQSATROEGNAWQ 1318
Qy 1367 RNEDGSLI-----INVQHTENRWAVTGRTEYDNKQPIRTYQYFLNDWRVYNSDSAR 1419
Dy 1319 RGRDGLVTASDGLPVTTA-TNFRWATGRAEYDNKGLPVRYQYQYFLDSHQYVSDSAR 1377
Qy 1420 QEKAYADTHYDPIGREIKVITAKGWRFTLFTFWFTVNEDENDT 1465
Dy 1378 Q--DLVADTHFYDPTAREWQVITAKGERQVLYTFWFVSEDENDT 1421

RESULT 15
ADR21540
ID ADR21540 standard; protein; 1428 AA.
XX
AC ADR21540;
XX
DT 04-NOV-2004 (first entry)
XX
DE Serratia SepB toxin SEQ ID NO:60.
XX
KW toxin; insect; insecticidal; transgenic; pest control.
XX
OS Serratia entomophila.
XX
PN WO2004067727-A2.
XX
PD 12-AUG-2004.
XX
PF 07-JAN-2004; 2004WO-US000394.
```

```
XX 21-JAN-2003; 2003US-0441723P.
XX (DOWC ) DOW AGROSCIENCES LLC.
XX
XX Hey TP, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
XX
XX Controlling or inhibiting an insect, useful for pest control, comprises
XX contacting the insect with effective amounts of a Protein A, a Protein B,
XX and a Protein C.
XX
XX Claim 1; SEQ ID NO 60; 368pp; English.
XX
XX The invention relates to a novel method for controlling or inhibiting an
XX insect comprising contacting the insect with effective amounts of a
XX Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
XX is encoded by a naturally occurring gene or has an amino acid sequence
XX that differs from the product encoded by a naturally occurring gene only
XX by truncation or by conservative amino acid changes. Protein A is a 230-
XX 290 kDa toxin complex insect toxin that is derived from a first taxonomic
XX species, has stand alone insecticidal activity, and has an amino acid
XX sequence at least 40% identical to a sequence selected from XptAlwI,
XX XptA2wI, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
XX complex potentiator having an amino acid sequence at least 40% identical
XX to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wI, XptB1xb,
XX PtcB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
XX having an amino acid sequence at least 35% identical to a sequence
XX selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptC1xb, PtcC1, PtcC1
XX (orf 6 long), PtcC1 (orf 6 short), and SepC. Also claimed is a transgenic
XX plant or plant cell that produces a Protein A, a Protein B, and a Protein
XX C. The method is useful for pest control. The present sequence represents
XX Serratia entomophila SepB toxin.
XX
XX Sequence 1428 AA;
XX
XX Query Match 50.3%; Score 3971; DB 8; Length 1428;
XX Best Local Similarity 52.7%; Pred. No. 2.3e-309;
XX Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;
XX
Qy 1 MONSQPSITELSLPGGGAITGMBALPTGPDGMAALSPLPISAGRYAPAFITLYN 60
Dy 1 MONHQDMAITAPTLPSCGGAVTGLKIDAAAGDGAATLSIPLSPSGRYAPATGALNYH 60
Qy 61 SGAGNSPFGGWCNVTIRRTTHFGVPHYDETDTLFGPEGEVLVA-----DQPDDES 114
Dy 61 SRSGNGFGGIGGGAAGVQRRTRNGAPTVDYDFTGPDGEVLVPALTAAQTQEARQAT 120
Qy 115 TLQGINLGATFTVTGYSRSLRSHFSRLEYWQPKTTGKTDFWLIYSPDQVHLLKSPQAR 174
Dy 121 SLGGINPGGSFNVQVRSRTEGSLRSLRLWLPADETETEFWLYTPDQVALLGRNAQAR 180
Qy 175 ISNPSOTTQTAQWLLSEASVSSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHYVY 234
Dy 181 ISNPTAPTQTAQWLLSEASVSSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHYVY 240
Qy 235 GNRITASETLPCLDGSAAPSQADWLPYLVFDYGERSNMLKTPPAFTTGS--WLCHQDFSR 292
Dy 241 GNRQAARTLPAL--VSTPMSDMSWLFILVFDYGERSSVLSSEAPAWQTGSGEMLCRQDCFSG 299
Qy 293 YEYGFETRRLCRQVLMVHHQLALDSKITHNGPTLVSRLLIINYDSSASTASTLVFVRV 352
Dy 300 YEFQFNLRRLCRQVLMVHHQLALDSKITHNGPTLVSRLLIINYDSSASTASTLVFVRV 359
Qy 353 GHEQDGNVVTLPPLLAYQDFSPRRHAWQPMVDLANFNANAIQRWQLVDLKGEGPLGLLYQ 412
Dy 360 AVESDGTSCALPALALGWQTFPTTILSAWQTRDDMGKLSLLQPYQLVDLNGEGVVGILYQ 419
Qy 413 DKGAWYRSQRLGETSGDAVTEKMQPLSVIISLQSNASLVDINGQGLDWTITGGLR 472
Dy 420 DSGAWMYRFPVRSQGDPTDAVTWGAAALPTMPALHNSGITLADNGDGLRLEWVVVTAFCVA 479
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:33:11 ; Search time 54 Seconds
(without alignments)
2626.361 Million cell updates/sec

Title: US-10-706-424-10
Perfect score: 7901
Sequence: 1 MQNSQPSITSLPKGGGA.....WFTVNEDENTAAEVKKVKM 1474

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.80.*

2: PIR1.*

3: PIR2.*

4: PIR3.*

5: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3569.5	45.2	1496	2 AH0447	insecticidal toxin
2	839	10.6	591	1 S09498	virulence-associat
3	835	10.6	591	1 S22664	virulence-associat
4	822.5	10.4	593	2 S15215	virulence-associat
5	709.5	9.0	563	2 S78224	virulence-associat
6	172.5	2.2	1404	2 E85509	hypothetical prote
7	172.5	2.2	1404	2 E30658	RhG core protein
8	164.5	2.1	1399	2 A99720	RhC core protein
9	161	2.0	1863	2 S46217	protein-tyrosine-p
10	160.5	2.0	1425	2 H4780	RhD protein precu
11	158	2.0	3083	2 A24393	hypothetical prote
12	156	2.0	2167	2 AF1499	cell wall-associat
13	155.5	2.0	1400	2 E30886	RhE core protein
14	153	1.9	3036	2 T18995	hypothetical prote
15	151	1.9	1377	2 E86034	RhA protein in rh
16	151	1.9	1394	2 H91236	RhA core protein
17	151	1.9	1409	2 F91187	RhA core protein
18	149	1.9	2314	2 T28698	hypothetical prote
19	149	1.9	1397	2 A85570	RhC protein in rh
20	149	1.9	1398	2 B85549	hypothetical prote
21	149	1.9	1907	2 S50893	protein-tyrosine-p
22	148	1.9	1398	2 H30698	RhD core protein
23	148	1.9	3972	2 S75251	hypothetical prote
24	147.5	1.9	682	2 C64898	RhE protein - Bsc
25	147.5	1.9	4199	2 S76412	hypothetical prote
26	145.5	1.8	632	2 F83387	copper resistance
27	144	1.8	709	2 S38241	hypothetical prote
28	144	1.8	1377	2 C65159	RhA protein precu
29	144	1.8	1411	2 E65145	RhB protein precu

30	142.5	1.8	1512	2 AH0439	probable membrane
31	141	1.8	1274	2 T37193	enamelin matrix pr
32	141	1.8	1397	2 C64805	RhC protein precu
33	141	1.8	2893	2 A64556	toxin-like outer m
34	139.5	1.8	1577	2 T30858	glucosyltransferas
35	139	1.8	828	2 G87584	hypothetical prote
36	139	1.8	2515	2 S47008	tenascin-like prot
37	138.5	1.8	1158	2 F90854	probable host spec
38	138	1.7	4427	2 FN0637	polyketide synthas
39	137.5	1.7	656	2 G85731	RhE element associ
40	137	1.7	2514	2 F81045	hemagglutinin/hemo
41	137	1.7	2628	2 T28651	hemagglutinin A -
42	136.5	1.7	1645	2 H85554	hypothetical prote
43	136.5	1.7	3566	1 A40701	tenascin-X precurs
44	136	1.7	775	2 C95921	hypothetical prote
45	136	1.7	1083	2 T23031	hypothetical prote

ALIGNMENTS

RESULT 1

AH0447

insecticidal toxin complex [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C;Accession: AH0447

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AH0447

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1496 <KUR>

A;Cross-references: UNIPROT:Q8ZAV4; UNIPARC:UPI000000CDA67; GB:AL590842; PIDN:CAC93148.1;

C;Genetics:

A;Gene: YPO3678

Query Match	45.2%	Score	3569.5	DB 2	Length	1496				
Best Local Similarity	49.3%	Pred. No.	3.8e-220							
Matches	748	Conservative	211	Mismatches	486	Indels	71	Gaps	29	
Qy	1	MONS-QDPSITSLPKGGG	AI	TGMBEALPTGPDGMAALSLPLPISAGRGYAPAF	T	LN	59			
Db	1	MENSKQAVAPLSLPKGGG	AI	TGMBEALPTGPDGMAALSLPLPISAGRGYAPAF	T	LN	59			
Qy	60	NSGAGNSPFLGMDCNVMT	IR	RRTHFGVPHYDETD	T	FLGPEGEVLVWADQPRD	--ESTLQ	117		
Db	61	SSGSGNGPFLGMDCNVMT	IR	RRTHFGVPHYDETD	T	FLGPEGEVLVWADQPRD	--ESTLQ	117		
Qy	118	GINLGATFTTGYRSRLES	HS	FLRYWQPKT	-	TGKTD	FWLIYS	PDGQVHLKGS	QPARIS	176
Db	121	SLN-GEQSVIRYLPRIE	GN	FRHRYEYRPR	T	NNNSQAPF	FWLHSSDQKHC	LGYSAAA	RIA	179
Qy	177	NPSQTTQAWLLASVS	SG	EQIYYVRAEDD	T	GCBADEI	----	THHLQATQRYL	HI	231
Db	180	DPLHPEHIAEWLLAES	VS	LSGHEHGYQ	QAEDEG	IDEPSIYKAEK	QNHPPAASQRYL	KR	239	
Qy	232	VYGNRTASSETLPGLD	GS	APSQADWL	FY	VPDYGER	NNKLT	PPAFSTTGS	WLCRQDRFS	291
Db	240	VYGNRQAAYELCLT	-	Q	PAPTS	WFLS	FLIFDHGEYS	NIABQVP	VI	298
Qy	292	RYEGFPIRTRRLCQV	LM	THLQALD	SK	ITEHNGP	TLVSR	LI	NYDESAT	351
Db	299	HPNYGFVTRRLCQV	LM	THLQALD	SK	ITEHNGP	TLVSR	LI	NYDESAT	351
Qy	352	VGHEODGNVVTLP	LE	LAYQD	FS	PRHHA	WQPM	DVLAN	FAIOR	411
Db	358	LAHEPDGTRKSLP	LE	PFYD	QD	FS	TRDAL	GWQ	PLTD	415

A;Molecule type: DNA
A;Residues: 1-563 <NOR>
A;Cross-references: UNIPARC:UPI000008BA96; EMBL:X57096; NID:g46998; PIDN:CAA40380.1; PID
R;Norel, F.; Pisano, M.R.; Nicoli, J.; Popoff, M.Y.
Res. Microbiol. 140, 455-457, 1989
A;Title: Nucleotide sequence of the plasmid-borne virulence gene mkb from Salmonella ty
A;Reference number: A43996; MUID:90161559; PMID:2696057
A;Accession: 823713
A;Molecule type: DNA
A;Residues: 1-513, 'NLYLKGKGRILGDVAHPKGEAELMF', 514-563 <NOW>
A;Cross-references: UNIPARC:UPI00001780F5; EMBL:X57096
C;Genetics:
A;Genome: plasmid
C;Superfamily: virulence-associated protein spvA
C;Keywords: DNA binding; transcription regulation

Query Match 9.0%; Score 709.5; DB 2; Length 563;
Best Local Similarity 39.1%; Pred. No. 1.7e-37;
Matches 170; Conservative 60; Mismatches 138; Indels 67; Gaps 13;

QY 9 ITLSLPKGGGALTGMGEALTPTGPGMAALSPLPISAGRYAPAPFTLYNAGNSPF 68
DB 15 ITPPFLPKG-----GKALSQSGPDGLASITLPLISAERGAPA----- 53
QY 69 GLGWDCNV-----MTIRRTTFG-----VPHYDETDFLGPEGLV-----VAD 108
DB 54 ---WRCTAAVAAMALRGVLVLRDNEHCPPHCPCVQYNDSDRFLGPDGEVLVQTLSTGD 110
QY 109 QPRDESTLQ--GINLGATFTVGYRSRLESHFRLEYWQPKTKTKTDFTWLIYSPDGOVHL 166
DB 111 APNPVTSFAYGDVSFPQSYVTYVQRTSSFFRYLEYWVGNNG--DDFWLLHDSNGILHL 169
QY 167 LKSPQARISNPQTTOAOWILLEASVSSRGEIYQYRAEDDTGCEADRIITHLQATAQ 226
DB 170 LGTAAARLSDPQAASHTAOWLEESVTPAGEHIYSYLAENGNDVNLNGNEAGRDSAM 229
QY 227 RYLHIVYGNRTASSETPLGIDGAPSOADMFLVDFYDGERSNLTKTPPAFTTGSMLCR 286
DB 230 RYLSKVQYGNATPAADL-YLWTSATPAVQMLFLVDFYDGERGVDPPVPAFTAQNSWLAR 288
QY 287 QDRFSRYEYFEINTRRLRQVLMYHLQALDKSKITEHNGFTLVSRILNYDSEASITL 346
DB 289 QDPFSLYNGFEIRLHRLCQVLMFHH---FPDELGE--ADTLVSRLLLEYDENPILTQL 343
QY 347 VFVRVCHQDG-----NVVTLPLLEYAQDFSPRHHAWQPMQMDVLANFNAI-----Q 394
DB 344 CAARTLAYEGDGYRRAPVNNMPPPPPPMMGGNSRSPKSKWAIVESKIQIALRYYSQAQ 403
QY 395 RWQLVD--LKGEGLP 407
DB 404 GYSVINKYLRGDDYP 418

RESULT 6
E85509
hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: E85509
R;Perne, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85509
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1404 <STO>
A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00000D025B; GB:AE005174; NID:g12512977; E
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0268
C;Superfamily: rhaF protein

Query Match 2.2%; Score 172.5; DB 2; Length 1404;
Best Local Similarity 19.4%; Pred. No. 0.021;
Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

QY 327 PTIVSRILIYNDESATIASTLVFVRRVGHQDGNVVTLPPLLEYAQDFSPRHH----- 379
DB 51 PVLGAKVLPGETDLALPGPLPFI-----LSRAYSSYRTTAPVGVFGP 94
QY 380 HMQ--PMDVLANFNAIQRMQLVLDKGBGLPGLLYQDKGAWWYRSARQLRGEIGSDAVTWK 438
DB 95 GWKAPFDI-----RLQIRD-----BGLILNDN-----GGRSIHPEPL 126
QY 439 QPLSVIPSLQSNASLVDINGDQGLDQVITGPGLRGVHSORP--DGSWTRFTPLNALPVEY- 496
DB 127 FPGEI--SYSRSESF-----WLARG-GVAEOHSSQPLSALW-----QVLPEVDR 167
QY 497 THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRDGF-----AK 534
DB 168 LSPHMYLATNSLQGPWWILNWPPEVFGADEVLPPPEPAYRVLTGVVDGFGRTLAFHRAAE 227
QY 535 GKDVVQSGDITLVPQADPKLVAFSDVLGSGQAHLEVV-----SATKVT--CHPNLGRGR 588
DB 228 G-DV--AGAVTGTVDGAGRR-----FHLVL--TTOAQAERVKORATSLSSPAGRSASS 279
QY 589 FGQPIITLPGFSQPATEF-----NPAQVYLADLDGSGPTDLIYVHTNRLDI 633
DB 280 LVPPDTLPA-----GTBYGADNGIRLEAVWLTHDPA--YPDELPAAPLARYTYTASSELRA 333
QY 634 PLNKSG---NGPABPVTLPFPEGLRFDHTCQLOMADVOGLGVASLILSVPHMSPHH----- 686
DB 334 VYDRSGTVRGFA-----YDAEHAG-----RMVAHVAGR 363
QY 687 ---WRCDLTNMPKPLNEMNNMNGVHTLRYRSSQFWMDEKAAALTTGQTPVCVLPFP 742
DB 364 PESRYDDTGR-----VTELNVNPEGLDYRFYEQ-----DRVTITDSLNRREVLY----- 409
QY 743 IHTLWQTEDETSIGNKLVLTLLRYARGAMDGRERFRGFGYVEQTD-----SHQLAQ 794
DB 410 -----TEGE--GGUKRVVKEHADGSTRSEYDEAG--RLKAQTDAAGRRTYSLHMAS 459
QY 795 GNAPERTPP-----ALTKNMYATGLFVIDNALSTEYMRDDQAFAGSPRFTT 841
DB 460 GAVTAVTGPDRTVRYGNSQRQVTSVTPDGL-----RSSREY----- 498
QY 842 WQDNKQVLTPEDDNSRYWFRNALKQOLLSELYGLDDSTNKIVPTVTEFRSQVRLOH 901
DB 499 -----DEKGLAAETSRSGETTR---YSYDDPASE-LPTGIQD-----A 533
QY 902 TDSRYVPLMSSVVESRNVHYERTASDPQCSNIT-LSSDRFGOLPKQLSVQYPRQCPAI 960
DB 534 TGS TKQWMS-----RYQLLITTDGSGYTRYERYDYGQ---QIAV-----HREGEI 578
QY 961 NLYPDTLPKLLANSYDDQORQLRLTYQSSWMHLLTNNTVRLVGLPDPSTRSDI--FTYG- 1017
DB 579 STYSYNPRQLVVSQKDAQGRETR--YEYSAAGDLT-----AIVAPDGSSEIYQDAWGK 631
QY 1018 AENVVAGGL-----NLELLSKNSLIADDK--PREYLCQOKTAYTDGQN----- 1059
DB 632 AVSTTQGLTRSMGYDAAGRITVLTNENGSGSTFRYPDPVDRLTEORG--FDGRTQRYHYD 689
QY 1060 -TTPLOTPTTQALIAFTTETTVFNQSTLSAPNGSIPSDKLSLTTLTEQAGYQOTNYL--PPRT 1116
DB 690 LTGKLQSEDEGLVTLWHYDASDRITHTRVNGD-----PAEQWQYDEHGWLTLTSH 741
QY 1117 GEDKVMVAHHGYTDYG-----TAAQFWRPQ-----KQSNTOJLTGK 1151
DB 742 SEGHRVSVHYGYDDKRLTGERQTVENPENGEMLEHETGHAYSEOGELATRQSPDGLPPV 801
QY 1152 ITLIWDANYC-----VVQTRD-----AAGLTTSKAYDWRFLTP 1185
DB 802 EWLTYGSGYLAKMGLGOTPLVYMRDLRHRETARSFGGEAYELATAMWNTSGQLRSRLNL 861

```
Qy 1186 VOLT---DINDNQHLLITLDALGRPITLRFWGTENGKMTGYSSPEKASFPSPDYNAAIEL 1242
Db 862 POLDRDYDNDNGQLIRISGPQSEYRY--SDTGLTGVHT-----TAANLDIDI 910
Qy 1243 KKPLPVAQCQVYAPESWMPVLSQKTFNRLAQDQWQKLYNARIITEDGRICTLAYRRWVQS 1302
Db 911 -----PYATDPAGNRLPD----- 923
Qy 1303 QKAIPQLISLLNNGPRLPHSLTLTTDRYDHPDPEQIRIQ---VVFSDGFGRLLOAAARH 1359
Db 924 -----PELHPDS-TLTA-----WPNRIAEADAHVYRVYDEVYGRLEAKTDRI 963
Qy 1360 EAGMARQBNEDGSLIINVQHTENRWATGRTYDYNKQPIRTYQPYPLNDWRYVNSDAR 1419
Db 964 PEGVIRMHDE-----RTHYHDSQHLVFTHTRIQ 993
Qy 1420 QEKEAYADTHVDPIGREIKVITAKG-WFRRTLFTPMFTVNEDENDT 1465
Db 994 HGEPOVESRYLYDPLGR-----TGKRVRRERDLTGWMSLSRKPEET 1036

RESULT 7
RhsG core protein with extension [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90658
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; PMID:21156231; PMID:11258796
A;Accession: E90658
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1404 <HAY>
A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BAB33660.1;
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: EC0237
C;Superfamily: rhaP protein

Query Match 2.2%; Score 172.5; DB 2; Length 1404;
Best Local Similarity 19.4%; Pred. No. 0.021;
Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

Qy 327 PTLVSRLLINVDSEAIASLTLPVRRVGHQGVNVTLPPLLEYADQFSRHH----- 379
Db 51 PVLGAKVLPGETDLALPGPLPFI-----LSRAYSYNTRTPAPVGVFGP 94
Qy 380 HWQ-PMDVLANFNATQRMQLVDLKGGLPGLLYQDKGAWYRSQRLGEIGSDAVTWKEM 438
Db 95 GWKAPFDI-----RLQIRD-----EGLILNDN-----GGRSIHFEPL 126
Qy 439 QPLSVPSLQNASLVDINGQGLDWVTGPGLRGYHSORP-DGSWTRFTPLNALPVEY- 496
Db 127 PFGBI--SYRSSESP-----WLARG-GVAEQHSSQPLSALW-----QVLPEDVR 167
Qy 497 THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRGF-----AK 534
Db 168 LSPHMYLATNSLQGPWMLNWPVRVPGADEVLPEPPPAYRVLVTGVVDGFGRTAFRAAE 227
Qy 535 GKDVQVSGDITLPVPGADPRKLVAFSDVLGSGQAHLEVE-----SATKVT--CWNLGRGR 588
Db 228 G-DV--AGAVTGVTDGAGR-----FHLVL-TTQRAEVFRKQRTSLSSPAGRSASS 279
Qy 589 FGQPIITLPFGSQATEF-----NPAQVYLADLQSGPTDLIYVHTNRLDI 633
Db 280 LVFPDTPA-----GTEYGADNGIRLEAVWLTHDPA--YPDELPAAPLARYTYTASGELRA 333
Qy 634 FLNKGSG---NGFAPVTLRFPFEGRLFDHTCQLQADVQGLGVASILLVSPHMSPH----- 686
Db 334 VYDRSGTQVRGFA-----YDAEHAG-----RMVAHHYAGR 363
```

```
Qy 687 -----WRCDLTMKPKWLLNENNNNGVHHTLRYSSSQFWLDEKAAALTTQTPVCYLPFP 742
Db 364 PESRYRYDDTGR-----VTELVNPEGLDYREYQG-----DRVTITDSLNRRELY- 409
Qy 743 IHTLWQTEDEISGNKLVTTTLVARGAWDGRERFPGVVEQTD-----SHOLAQ 794
Db 410 -----TEGE--GGLKRVVKKHADGSIITRSEYDEAG-RLKQCTDAAGRTESYSLHMAS 459
Qy 795 GNAPERTPP-----ALTKNWTATGLPVIDNALSTEYWRDQAFAGSPRFTT 841
Db 460 GAVTAVTGPDRTVRVGYNSQRQVTSVTPDGL-----RSSREY----- 498
Qy 842 WQONKOVPLTPEDDNRSYFNRLKQGLLSELYGLDDSTNKHVYTVTTPRQVRRLQH 901
Db 499 -----DEKGLAAETSRSGETTR---YSYDDPASE-LPTGIQD-----A 533
Qy 902 TDSRYPVLWSSVVSERNYHYERIASDPQCSNIT-LSSDRFGOPKLSVQYPRQOQPAI 960
Db 534 TGSTKQMAWS-----RYGQLLTFDSCGYTRYRYDRYQ--QIAV-----HREEGI 578
Qy 961 NLYPDTLPDKLANSYDDQORQLRLTYQSSSWHLLTNNTVRLGLPDSRSDI--FTYG- 1017
Db 579 STYSSYNPRQLVSKDAQGRETR--YEYSAAGDLT-----AIVAPDGSSEIQYDANGK 631
Qy 1018 AENVPAAGL-----NLELLSDKNLSIADDK--PREYLOQOQKTAITDQON----- 1059
Db 632 AVSTTQGLTRSMGYDAAGRITVLTNENGSQSTFRYPDVRDLTEORG--PDGRTQRYHD 689
Qy 1060 -TPTLQTPTRQALIAFTETTVFNQSTLSAFNGSIPSDKSLTLEOAGYQQTNYL--FPRT 1116
Db 690 LTGKLTQSEDEGLVTLWHYDASDRITHTVTNGD-----PABQWQYDSHGLWTLTSLHT 741
Qy 1117 GEDKVVVAHGYTDYG-----TAAQFWRPQ-----KQSNLTQLTGK 1151
Db 742 SEGHRVSVHYGYDDKGELTGERQTVENPETGEMLWEHETGHAYSEQGLATRQEPDGLPPV 801
Qy 1152 ITLIWDANYC-----VVQTRD-----AAGLTTSAKYDWRFLTP 1185
Db 802 EWLTYGSGYLAKMKLGCTPLVEYMRDLRHRRETARSFGGEAYELATAMWTSQLASRHLN 861
Qy 1186 VOLT---DINDNQHLLITLDALGRPITLRFWGTENGKMTGYSSPEKASFPSPDYNAAIEL 1242
Db 862 POLDRDYDNDNGQLIRISGPQSEYRY--SDTGLTGVHT-----TAANLDIDI 910
Qy 1243 KKPLPVAQCQVYAPESWMPVLSQKTFNRLAQDQWQKLYNARIITEDGRICTLAYRRWVQS 1302
Db 911 -----PYATDPAGNRLPD----- 923
Qy 1303 QKAIPQLISLLNNGPRLPHSLTLTTDRYDHPDPEQIRIQ---VVFSDGFGRLLOAAARH 1359
Db 924 -----PELHPDS-TLTA-----WPNRIAEADAHVYRVYDEVYGRLEAKTDRI 963
Qy 1360 EAGMARQBNEDGSLIINVQHTENRWATGRTYDYNKQPIRTYQPYPLNDWRYVNSDAR 1419
Db 964 PEGVIRMHDE-----RTHYHDSQHLVFTHTRIQ 993
Qy 1420 QEKEAYADTHVDPIGREIKVITAKG-WFRRTLFTPMFTVNEDENDT 1465
Db 994 HGEPOVESRYLYDPLGR-----TGKRVRRERDLTGWMSLSRKPEET 1036
```

RESULT 8

A99720
RhsC core protein with extension [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 31-Dec-2004
C;Accession: A99720
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; PMID:21156231; PMID:11258796

Qy 1421 EKEAYADTH-VYDPI-GREI 1438
Db 1210 ESGLYNHRHYDPLQGRYI 1229

RESULT 11
AH2493
hypothetical protein all7128 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AH2493
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2493
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3083 <KUR>
A:Cross-references: UNIPROT:Q8YL10; UNIPARC:UPI000000CEEDB; GB:BA000020; PIDN:BA078212.1;
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all7128
A:Genome: plasmid

Query Match 2.0%; Score 158; DB 2; Length 3083;
Best Local Similarity 18.8%; Pred. No. 0.66;
Matches 276; Conservative 163; Mismatches 507; Indels 520; Gaps 73;

Qy 1 MONSDPSITSLSPKGGGATGMEAL-TPTGPDG-----MAALSPLPISAGRGA 52
Db 725 IEAAAEVNVAASAGAGGGIIGTGLNLDKPTPGDKVRGNEFVQLLNPIEMFDASGLV 784

Qy 53 PAPTLYNSGAGN-----SPFG-----LQWDCNVWIR----- 80
Db 785 QAYLMAYAKVAGKVVKRIESPVKLLPGYKVBETPPOLHLATDIGGNLRNMGPNAAA 844

Qy 81 -----RRTHFGV-----PHYBTD 94
Db 845 REIINTEDGAEVFTTDDGKLTVAFNIPTQYSGVSKIIADGGTNDTIEIKDIEISA 904

Qy 95 TPLGPEGLVAVDQPRDESTLOGINLGAFTVTGYRSLRSHFSLRLEWQPKTKGTDF 154
Db 905 DLKGAGEDLYCGGSD--TIRG--GAD-----WDRLYGDRDD 940

Qy 155 WLIYSPDQVHLGKSPQARISNPSQTTQTAQWLLASVSRGQIYYQYRAEDDTGCEA 214
Db 941 P-VYGDGDDWLDG-GAGADTLNGAGPDT-----ASYTSATSAISNLVTQVSTGDAA 992

Qy 215 DEITHHL-QATAQRYLHTV-----YGNRTASSETLPG-----LDG--- 248
Db 993 DDVQSQTEQIVGSRYYDTLIGDEBNNEPDCGEGNDPISGGAGDRLSPGWDGDDVIDGTG 1052

Qy 249 -----SAPQA-----DMLFYLVDYGERSNLKTTPAFSTGWSLGRQD 288
Db 1053 TDTLVIDYSSLPTQAVAWSELDPNTSDWFFVAVAYGIGA-PIKTD--INVSGN----- 1103

Qy 289 RFSRYEYGFETRRLCRQVLMYHHLQALDSKITEHNGPTLVSLRLNLYDESAITLVF 348
Db 1104 -----YHATLSAD-----GLTVAGSGILGNSGS--GNOGLW 1132

Qy 349 VRRVHEODGNVVTLPPELAYQDFSPRRHAH--WQPMDV--LANFNAIQWQLVDLK-- 402
Db 1133 VKKI-HSSDPVAVIPNNQV-YQPLLEDGSKVWVSQDSDTWIANTGTQVRQTKLSIN 1190

Qy 403 -----GEGPLGLLYQDKGAWWYRSAQRIGEI-----GSDAVTWKRNQPLSPISLQSN 450
Db 1191 IGYGDGDLATISDGGSTIAMLSKRNNDKFTYTFIFIANADGKNLRQIN-----IPTSGG 1246

Qy 451 ASLVDINGDG-QLDWVITGPGRLGVRHSQRPGSWTRFTPLNALPVETHPRQAADLMGA 509

Db 1247 VRELDLSADGSKITWSQDG---GY---GPGGVWVANT---DGTNIRELSGNLYG 1292
Qy 510 GLSDLVL-----IGPKSVRLYANTRDG--FAKGQVVQSGDITLTPVPGADPR 554
Db 1293 NINPSISADGSTVVMAGYQAGYASTNLXAATDGSRFVVPVNTVEEVEGFAQAQSLAGDSR 1352

Qy 555 KLV-----AFSDVLGSGQAHILVEVSATKVTCTWPNLGRGRFGQPTLTPGFSQAT 603
Db 1353 RVVFTKFNQSDYSLYVGDIDGIEPQILIDASS-----PNIGIRGRHALSSYVDLVGRYN 1406

Qy 604 EKNPA-----QVYLADLDSGSGPTDLIYVHTRLDIF-----LNKS--GNGPAEP 645
Db 1407 SFDPATSGSIYT-----WGSRIYRNSNFERFDIIGTRYGDELPGNLDLDSLMGGGAD- 1460

Qy 646 VTLRFPEGLRFDHTCQLQMAADVQGLGVASLILSVPHMSPHHWRCDLNTMKPWLINEMNN 705
Db 1461 -TLK--AGLGDD---IYILDQWAGSQIE-----DAGGTD--TLRLITRN 1498

Qy 706 MGVHHTLYRSSSQFWLDEKAAALTTGQTPVCYLPFPIHTLMQET-----EDEISGN 758
Db 1499 PGATNTPRITDA-----DLSLAVPTTG-----IFGMRRAAGTSLIIDLNKDGIAAS 1543

Qy 759 KLV-VTTLRYARGAWDGRERFRGFGYVEQTDHQLAQGNAPERTPPALTKNWTATGLPVI 817
Db 1544 KTDLTILNFPDVTGTG-----AGTGFLETVANLAGAB-----ILSK-----LQVG 1583

Qy 818 DNALS-----TEYWRDDQAFAGFSRFTTW--QDNKDVPLTPEDDNSRY----- 859
Db 1584 DDTISGSAADDFDGLWLSNDTLSCGAGNDTLWGQDGNDF-LNGEDGNDLSQGGNGNDTLT 1642

Qy 860 --WPNRAKQLRLSELYGLDDSTNKHVPYVTTFRSQVRLQHTSDRYPVLWSSVRESR 917
Db 1643 PGWGNVDVDDG-AGTDVLVD-----YSNLNTRA-----VAMRTLSGTS 1680

Qy 918 NYHYERTASDPQCSQNTLSSDRFGQPKQLSVQYPRQQAIPNLYDTPDLKLLANSYD 977
Db 1681 GNYLQKE-----FIGNAYGLGTPLKIRET-----NSVSDKFALSA-- 1715

Qy 978 DQORQLRLTYQSSWHHLTNTNVRVLGLPDSRSDIFTYGAENVPAAGLNLLELSDKNSL 1037
Db 1716 -----DGTTVAYTYINNDPANGMLWIKKIDDSGGL 1746

Qy 1038 IADDKPREYLG---OQKTAFTDQNTTPIQTPTRQALIAFTETTVFNQSTLSAFNGSIP 1093
Db 1747 VKIDEIATEIALSDTGKIAWSGWN-----RVVAVANTNGTEKIRINLANNINGYI 1796

Qy 1094 SDKLS-----TTLBQAGYQQTNYLFPRTGEDKVVVAHHGYTDYG 1132
Db 1797 SLSLSGDGSQVSWNNGNQLLVANTDGTNIREITQSSSTKSLSENGSQIHWAGYQG-EKYG 1855

Qy 1133 -----TAAQFWRPQKSNQNTLTGK--ITLIWDANCYVVVQTRDAAGLTTSAK-YDWRF 1182
Db 1856 IWSNSTSTSLPVVKSVDGNLSLSSSDCIKAIWQDRYFLSVSSTNSTEIQOVAESYDFRV 1915

Qy 1183 L-----TPVQLTD-----IN-DNQ 1195
Db 1916 VGSSEPVLAADGAKVAFIKAINADNQ 1941

RESULT 12

AF1489

cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - L

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AF1489

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

Db 276 LSDSAPPDTLP-----TEYGFDRGIRLSAVLTHDPAYPESLPQAPLARYTYTEAGELL 330

Qy 633 IFLNKS-----GNGFAEPTVTLRPFPEGLRFDHTCQLQADVQGLGVAS 674

Db 331 AVYDRSNTQVRAFTYDAQHGRMVAHRVAGRPEMKY-----RYDDTGRV-----VEQLNPAG 382

Qy 675 LILSVPHMSPHHRCDLTNMKPWLNNNNMGVHHTLR--YRSSQFWLDEKAAALTG 732

Db 383 LSVRYQEQDIRITVTDLSLRREVLTHT--GGAGLKVVKKSLADGVSVTHSGYDAAGRLTA 440

Qy 733 QTPVCYLPFPPIHTLWQTEDEISGNKLVTLRVARG-----AWDGREREF----- 778

Db 441 QT-----DAAGRTYEGLNVVSGDITDITTPDGRTEKYNDGNQL 481

Qy 779 -----RFG-----YVEQT-----DSHQLAQGNAPERTPPALTKN-----W 809

Db 482 TAVVSPDGLSRRAYDEPGLRVSETSRCGDIVRAYDN-PHSELPAITTDATGSTRQMTW 540

Qy 810 --YATGLPVID-NALST--EYWRDDQAF-----GFSRPTTWQDNKDVLTPED---DN 856

Db 541 SRYGQLLAFTDCSGYQTRYEYDRFGQMTAVHREEGIS--RYRRY-DNRGLTSVKDAQGHE 598

Qy 857 SRYWFNRA-----LKGOLLRSELYGL-----DSD 880

Db 599 TRYEYNAAGDLTAVITPDGRSETQYDAGKAVSTTQGLTRSMYDLAGRITLTHENG 658

Qy 881 TNKHVYTVTFERSQVRLQHTDSRYPVLMSSVSVESRNYHYERIASDPQCSQN---ITL-- 936

Db 659 SRSEFTYDALDRLVQORGF--DGR-----TQRYHYDLTGKLTQSEDEGLVTLWH 705

Qy 937 --SSDRP-----CQPLKQLSVQVPRR-----QPAINLPDPTLPDKLLANSY 976

Db 706 YDESRLTRTVNGEPAEQW--QYDEHGLWTEIHLSEGHQVAVHY-----GY 751

Qy 977 DQOQRL--RLTYQOSS-----WHLTNNTVRLVGL-----PDS-TRSDIFTYGAENVPA 1023

Db 752 DDKGLAGEROTVINPETGELLWQHEHAHYNEQGLANRVTPDLSRVEWLTYS----- 806

Qy 1024 GGLNLELLSDKNLIADDPREYLGQOKTAYTDGQNTPLTPTROALIAFTETTVFNQS 1083

Db 807 -----GYLAGMKL-----GGTFLVEFTDRL-----HRE 830

Qy 1084 TLSAFNGSIPSDKSLSTLEQAGYQTNVLPRTGEDKVVAAHGYDYGTAQFWRPQK- 1142

Db 831 TVRSFGNN--AYELTSTYPAGLQSQRL-----NSQYDRDYDNONGDLVRSGRQT 883

Qy 1143 -QSNTQLTGKI---TLIWDANYCVVQTRDAAG-----LITSA 1176

Db 884 WEYGSATGRLESVRTLASDLDIRIPYAT-DPAGNRLPDPDLHPDSTLTWAMPDNRIADA 942

Qy 1177 KYDRFLTPVQLTDIN-----DNQHLITL----- 1200

Db 943 HYVVRHDEYGRLEKTDRI PAGVIRTDDETRTHYHYSQHLRVFYTRIQHGEPLVESRYL 1002

Qy 1201 -DALGRPITLFPWGTENGKMTGYS-----PKASFSPSDVNAIELKPLPVAQCO-VY 1254

Db 1003 YDPLGRMAKVRWRER-DLTGWSLSRKPEVTWYGMGDRLTTVQ-----TDTTRIQTVY 1057

Qy 1255 APESMMPVLSOKTNRLAEOQWOKLYNARIITEGRICTLAYRRVWQSOKAIPQLISLN 1314

Db 1058 EPGSFPLIRIVETENGEREKA-QRRSLAETLQOEGS-----E 1093

Qy 1315 NGPRLP-PHSLTLTTDRYDHPDQOIQOQVFSFGFGRLLQAAARHAGMARQNE--G 1371

Db 1094 NGHGVVPAELVRLDL-----EHEIRADRVSSSRRAWLAQGLTVEQ-LARQVEPEYTP 1148

Qy 1372 SLIINVQHTNR-----WAVTGRTEYDNKGQPIRTYQPYFLND-WRYVSND 1416

Db 1149 ARKVHFVHCDHRGLPLALISBDGNTAW---RGEYDEMGNLNENPYLHQPVRLLPQQ 1204

Qy 1417 SARQEKAYATHVVDPI-GREI---KVITAKGW 1446

Db 1205 HDEESGLYNNRYDPLQGRYIITQDPIGLAGGW 1238

RESULT 14

T18995

hypothetical protein C06B8.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18995

R:Steward, C.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19057

A:Accession: T18995

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3036 <WIL>

A:Cross-references: UNIPROT:O17575; UNIPARC:UPI000017B740; EMBL:Z81463; PIDN:CAB03852.1;

C:Experimental source: clone C06B8

C:Genetics:

A:Gene: CESP:C06B8.7

A:Map position: 5

A:Introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/3

Query Match 1.9%; Score 153; DB 2; Length 3036;

Best Local Similarity 18.7%; Pred. No. 1.3;

Matches 241; Conservative 164; Mismatches 459; Indels 428; Gaps 65;

Qy 910 ---WGSVBSRNYHYERIAS--DPCSONITLSSDRFGQLKLSVYQPRROCPAINLY 963
Db : : : : :
711 ILISWSVEFLKNQVHPSMEIFSCDPDVINIV--DVTGNRIDG-NLGYGMRAPAVNMH 766
Qy 964 PDTLPDKLLANSVDDQQRQLRLTYQQSSWHHLTNVTYVLGLPDSRTSDIFTYG-ANVP 1022
Db : : : : :
767 TVINSNQFLNN-----DTTLIRNAQNPELDPAEV-----TISKNVKFENVAKYII 815
Qy 1023 AGGLNLELLSDKNSLIADDKPREVL--GOQ-----KTAYTDGONTTPLOQTPTQAALIAFT 1075
Db : : : : :
816 SIGN-----EDAPRGFLTFNQNEIRANTVFDPFSLPRSTPYAALVVSSS 863
Qy 1076 ETTVFNGSTLSAFNGSPSKLSTLEQA-----GYOQTYNLFEPRTGEDKVWVA 1124
Db : : : : :
864 NVKIHR---NCFNNERSKYEIATELEQHAKWIDARENWGFQE---IPRF-IDKFF-- 912
Qy 1125 HHGYTDYGTAAQFWRPQKQNTQLTGITILIMANY-----CVVOTRAAGL 1172
Db : : : : :
913 -DOFNRYSLASIDIEPTYMAACNORMPYITLLNGAFRQFKQSBPFKLGGIYYENHDLLKG 971
Qy 1173 TTSKVDWRFLTVPQVLT----DINDNOHLITLDAIGRPITLRFWGTENGHWTCYSSP--- 1225
Db : : : : :
972 RYQVTEDLQVVPQAKLTIASGVLEFOHGIMLVQGDLIRNEY--DODEKVIETSTPFTL 1029
Qy 1226 EKASFSPSPDVNAAILK-----KPLPVAQCQVYAPESWMEVLKSQTFNRILA-BQDWOKLY 1280
Db : : : : :
1030 ERRSDIRLVDGNYEYTEGLEVLVDQGWTCNRSWTQTLTILACNLGLVADVQYPE 1089
Qy 1281 NARIITEDGRICTLAYRRWWQSOKAIPOLISLANNRPLPHSLTLTTDYRDHPBQQIR 1340
Db : : : : :
1090 NWRIFFPGD-----LPMIMDNIRCENEVEDITKCRH----- 1121
Qy 1341 QOVVFSDFGRLQLQAAARHAGMARQENEGSLIINVQHTENWA-----VTG 1388
Db : : : : :
1122 -----DGVERNCAGCR-----STEVGRLCULEPRWAGRVYSLLANPPVTG 1163
Qy 1389 RTEYDN-----KGQ--PIRT--YOPYFLNDWRY 1412
Db : : : : :
1164 OTTMNDNWRIEKGLFNFRTESEFCFAFKVDWNY 1195

RESULT 15
E86034
rhaA protein in this element [imported] - Escherichia coli (strain O157:H7, substrain EDL)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: E86034
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001.
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: E86034
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1377 <STO>
A/Cross-references: UNIPROT:Q8X418; UNIPARC:UPI000000DODCF; GB:AEO05174; NID:g12518335; F
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: rhaA
C/Superfamily: rhaF protein

Query Match 1.9%; Score 151; DB 2; Length 1377;
Best Local Similarity 18.5%; Pred.No.0.48;
Matches 227; Conservative 136; Mismatches 413; Indels 448; Gaps 63;

Qy 406 LPGLLYQDKGAWWWRSQAORLGEISGDVATWEKKMQLSVIPSLOSNASLVINGDGOLDMW 465
Db : : : : :
70 LPFILSTYSYRXTKTPAPVSGISGP---WK--MPADIRQLRDNTLILSDNGRSLYFE 124
Qy 466 ITOPGLRGYHSRQPDGSGWT-----RFTPL-NALPVRY-THPRQAQLAD----- 505

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:37:07 ; Search time 13 Seconds
(without alignments)
763.846 Million cell updates/sec

Title: US-10-706-424-10
Perfect score: 7901
Sequence: 1 MQNSQPSITELSLPKGGGA.....WFTWNEDENTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 51470 seqs, 6736768 residues

Total number of hits satisfying chosen parameters: 51470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New*

1: /cgm2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
2: /cgm2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
3: /cgm2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
4: /cgm2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
5: /cgm2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
6: /cgm2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
7: /cgm2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep.*
8: /cgm2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6031.5	76.3	1476	6	US-10-647-956A-4
2	139	1.8	2515	7	US-11-113-424-53
3	136	1.7	2769	7	US-11-113-424-14
4	119	1.5	1227	6	US-10-793-626-96
5	116	1.5	1348	6	US-10-995-561-624
6	116	1.5	2715	7	US-11-113-424-51
7	115.5	1.5	1192	6	US-10-858-730-72
8	114.5	1.4	1107	6	US-10-485-517-145
9	114.5	1.4	4419	6	US-10-821-234-1155
10	114	1.4	805	6	US-10-485-517-198
11	113.5	1.4	1206	6	US-10-858-730-73
12	113.5	1.4	4128	6	US-10-770-726-77
13	113	1.4	1907	7	US-11-000-463-250
14	112.5	1.4	915	6	US-10-647-956A-6
15	112.5	1.4	1432	6	US-10-510-386-218
16	112	1.4	2516	6	US-10-647-956A-2
17	110	1.4	2296	6	US-10-995-561-633
18	110	1.4	2355	6	US-10-995-561-623
19	110	1.4	2355	6	US-10-995-561-627
20	110	1.4	2384	6	US-10-821-234-1545
21	110	1.4	2386	6	US-10-995-561-626
22	108.5	1.4	782	6	US-10-972-053-2
23	108.5	1.4	782	6	US-10-972-053-8
24	108.5	1.4	1897	6	US-10-821-234-1635
25	107.5	1.4	1315	6	US-10-995-561-630

ALIGNMENTS

RESULT 1

US-10-647-956A-4
; Sequence 4, Application US/10647956A
; Publication No. US20050251878A1

; GENERAL INFORMATION: Constant, Richard

; APPLICANT: Bowen, David

; APPLICANT: Rocheleau, Thomas

; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: 61645

; CURRENT APPLICATION NUMBER: US/10/647,956A

; CURRENT FILING DATE: 2003-08-26

; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514

; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: US 60/191806

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 4

; LENGTH: 1476

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-10-647-956A-4

Query Match 76.3%; Score 6031.5; DB 6; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Qy 1 MQNSQPSITELSLPKGGGAITGGEALTPGDPGMAALSPLPISAGRYAPFTLYN 60

Db 1 MQNSQPSITELSLPKGGGAITGGEALTPGDPGMAALSPLPISAGRYAPFTLYN 60

Qy 61 SGAGNSPFGCGWDCNVMTIRRTFHGVPYHDETDFTLGPGEVLVA-----DQPRDES 114

Db 61 SGAGNSPFGCGWDCNVMTIRRTFHGVPYHDETDFTLGPGEVLVA-----DQPRDES 120

Qy 115 TLQINLGATFTVTGVRSLRSHFSRLRYWQPKTTGKTDFWLIYSPDQVHLLGKSPQAR 174

Db 115 TLQINLGATFTVTGVRSLRSHFSRLRYWQPKTTGKTDFWLIYSPDQVHLLGKSPQAR 180

Qy 175 ISNPSQTTQTAQMLLEASVSRGEIYYQVRAEDDTTCRADEITHHLOAQAQRYLHVY 234

Db 175 ISNPSQTTQTAQMLLEASVSRGEIYYQVRAEDDTTCRADEITHHLOAQAQRYLHVY 240

Qy 235 GNTASTETPLDGSAPSOADWLFYLYFDYGERSNLKTTPASTTGTGSLCQRDFRYE 294

Db 235 GNTASTETPLDGSAPSOADWLFYLYFDYGERSNLKTTPASTTGTGSLCQRDFRYE 300

Qy 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Db 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Qy 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Db 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Qy 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Db 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Qy 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Db 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Qy 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

Db 241 GNLTAADVFTTLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRDRFRYE

```
QY 295 YGFEIRTRLCROVLAMVHHLQALDSKITTEHGPTLVSRLLNLYDESAIASTLTVFVRVGH 354
Db 301 YGFELTRRLCROILMFHRLQTLUSGAKGDDPEALVSRLLIDYENAMVSTLVSRRVGH 360
QY 355 EODGNVTVLPPLBELAYQDFSPRHIAHQWMDVLNANFNAIQRMQLVDLKGEGPLGLLYQDK 414
Db 361 EDNNTVTALPPLBELAYQDFPEQETALWQSDVLNANFNTIQRMQLLDLKGEGVPGILYQDR 420
QY 415 GAWWYSARLGBIGSDATWEMKQPLSVIPSLQSNASLVINDINGDQOLDWVITPGPLRGY 474
Db 421 NGWYYSARQAGEENNAVWGMQQLLPITPAVDNASLMDINDGQOLDWVITPGPLRGY 480
QY 475 HSORPDGSTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRGEPK 534
Db 481 HSQHPDGSWTRFTPLHALPIEYSHPRQAQLADLMGAGLSDDLVLGPKSVRLYVNNRQDFT 540
QY 535 GKDVQSGDITLTPVGADPRKLVAFSDVLGSGQAHLEVSATKVTWCPNLRGRFQOPIIT 594
Db 541 GRDVQSGDITLTPVGADARKLVAFSDVLGSGQAHLEVSATQVTCWPNLHGCRFQOPIV 600
QY 595 LPQFSQATEFNPAQVILADLDSGPTDLIYVHTNRDLIPFKSGNGFAPBPVTLRPPGL 654
Db 601 LPQFSQAASFNDRVHLADLDSGPDALIYVHADRLDIPFSNESGNGFAPKPTLSPPDGL 660
QY 655 RPDHTCOLQWADVQGLGVASLIISVPHMSPHWRCDLTNMKPMLLNEMNNMGVHHTLRY 714
Db 661 RPDHTCOLQWADVQGLGVASLIISVPHMAPHHWRCDLTNAKPWLLSETNNMGANHTLHY 720
QY 715 RSSQFWLDEKAAALATGTGTPVCLPFPPIHTLMQTEDEISGNKLVTLRYARGAWDGR 774
Db 721 RSSVQFWLDEKAAALATGTGTPVCLPFPVHTLMQTEDEISGNKLVTLRYARGAWDGR 780
QY 775 EREFRGFGYEQTDSHQLAQNAPERTPPALTQNTATGLPVDNALSTBYWR-DQOAPA 833
Db 781 EREFRGFGYEQTDSHQLAQNAPERTPPALTQNTATGLPVDNALSTBYWRGDKQAPA 840
QY 834 GPSPTFTWQDNKVPITPDDNSRYWFNPAKQGLLSRLYGLDSTNKHVYVYVTEPR 893
Db 841 GPTFTFTLWKEGKDVLPITPDDHNLWNLNKAQGLPRLSELYGLDGSAAQOQIPIYVTESR 900
QY 894 SQVRLQHTDSRYFVLWSSVYSESNHYERIASDPQCSQNTILSSDRFGQPLKQLSVQYP 953
Db 901 PQVRLQDQATVPVLWASVYSESRVYERIIISDPQCNQDITLSSDLFGQPLKQSVQYP 960
QY 954 RROQPAINLYPDTLPDKLLANSYDDQORQLRLTYQSSWHLLTNTVTVLGLPDSRSDI 1013
Db 961 RRNKPTTNPYPDTLPDTLPASSYDDQQLRLTCROSSWHLLIGNELRVLGLPDRSDA 1020
QY 1014 FTYGAENVPAAGLNLELSDKNSLIADDKPREVLGQKATYATDQONTTLPQTPTROALIA 1073
Db 1021 FTYDAQVPVDGNLETLCAENSLIADDKPREVLNQORTYTTDQKQTPUKTPTROALIA 1080
QY 1074 FTETTVNQSTLAFNGSIPSDKLSTTLEQAGYQQTNYLPPRTGDKVWVAHHGYTDYGT 1133
Db 1081 FTETAVLTESLLGAFDGGITPDELPGLLTQAGYQQEPYLPRTGENKVVWVARQGYTDYGT 1140
QY 1134 AAFWRPQKQSTQGLTKITLIIDANYCVVVQTRDAAGLTTSKADNWRFTTPVQLTDIND 1193
Db 1141 EAQFWRPVPAQRNSLLTGKMTLKWDTYCVITQTDAAGLTVSANYDWRFTLPTQLTDIND 1200
QY 1194 NOHLITLDALGRPTLRFWCTENGKMTGYSSPEKASFSPSDVYNAATELKKPLPVAQCQV 1253
Db 1201 NVHLITLDALGRVPTQRFWGESVATGYSSSEKPSPPNDIDTAINLGPLFVAQCIV 1260
QY 1254 YAPESWMPVLQKTFNRLAEQDMQKLYNARIITEDGRICTLAYRRWVQSKALPQALISLL 1313
Db 1261 YAPDSWMPVLSQETFTNTLTQEEQETLDRSRIITEDWRIICALTRRWILOSQKISTPLVKLL 1320
QY 1314 NNGPRLPPHSLTTLTDYRDHDPQOQIRQOVVFSDFGFORLLQAAARHAGMARQORNDGSL 1373
Db 1321 TNSIGLPPHNLTLTLDYRDSDSEQIRQOVAFSDGFORLLQASVRHAGEAWORNDGSL 1380
```

```
QY 1374 IINVOHTENRMAVGTREYDNKGQPIRTYQPYFLNDRVYNSDSARQEKEAVADTHVYDP 1433
Db 1381 VTKVENTKRMVATGETEYDNKGQPIRTYQPYFLNDRVYNSDSAR--KEAVADTHVYDP 1438
QY 1434 IGREIKVITAKQWFRRTLTFTPMFTVNEDENDTAAE 1468
Db 1439 IGREIRVITAKWLRQSQYFPNFTVSEDENDTAAAD 1473

RESULT 2
US-11-113-424-53
; Sequence 53, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 2515
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-11-113-424-53
```

Query Match 1.8%; Score 139; DB 7; Length 2515;
Best Local Similarity 16.5%; Pred. No. 0.082;
Matches 236; Conservative 161; Mismatches 446; Indels 586; Gaps 59;

```
QY 1 MONSDPFSITELSLPKGGGAIITGMGEALPTGP----DGMALSLPLPISAGRGYAPFT 56
Db 1027 VRDNDYISQPELNWE----AVVSGRCLPGDEAHGCGDALAKADKALVPKGIATSSDNI 1082
QY 57 LNTNSGA-----GNSPPGLGW-----DCNVMTIRRRTHFGVPHYDE 92
Db 1083 LYFADCTNIRWVDRDGVSTLIGNHMKSHWKPIPCEGTLKLEEMHLRWPTELAVSPMDN 1142
QY 93 TDTFLG-----PGEVLVWADQPRDESTLOGINGATFTVTGYRSRLSEHSFRLSEYW 144
Db 1143 TLHIIDHMLRLMTPDGRVRSISGRP-----LHCATATAYDITDLATHTLV--- 1189
QY 145 QPKTKCTKDFWLIYSPDQVHLLGKSPQARISNPSQTTQTAQWLLSEASVSSRGEIYYQY 204
Db 1190 MPQS-----TAFGPGLELY-VAESDSQRINKRVVIGDRIAPFAGAESKNCNL---- 1237
QY 205 RABDDTGCEADEITTHLQATAQ-----RYLHVYVYGNRTASSETLPGLDGSAPS 252
Db 1238 -----ERGCDFEAEHYLATSAKENTIAALAVTPDSHVHIADQANYRIRSVMSSIPASPS 1293
QY 253 -----QADWLP-----YLVFDYGRSNNLKTTPAPFTTGSWLC 285
Db 1294 REYEIYAPDMQOEIYIFNRFQGHVSTRNLTGETTYFTYVNTVNTSNGKLSLTVTDAAGNKVF 1353
```



```
QY 841 TWQDN---KDVPLTPEDNSRYWFRNALKQGLLSRLYGLDLS-TNKHVPTVT--EPRS 894
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
776 PKEKTPMKEINLAP--DSSVVVUGLWATKYEVSYALKDILTSPQAQGVVTTLENVS 833
QY 895 QVRLQHTDSRYVPLWSSVVSRYHYERIASPQCSQNTLSDSRFGQPLKOLSVOYPR 954
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
834 PPRARVTD--TETTTIS-----WRT 854
QY 955 RQCPAINLYPDTLPDKLLANSYDDQORQLR--LTWCQSSWHHLTNTVRLGLPDSTRS 1011
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
855 KTIETIGFQVDVAP--ANGQPTIKIPDVRSVYITIGLQPGTDYKIYLYTLNDNARS 910
QY 1012 DIPTYGAENVFAGGLNLELL-SDKNSLIAD-----DKPREYL----- 1047
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
911 SPVVIDAATAIDAPSNLRFATTPNSLLVSMQPPRARITVYIKEYKPGSPREVPVPR 970
QY 1048 -----GQKATYAT--DGQNTTPL-----QTTRQALIAFTTTFVFNOSTLS 1086
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
971 PGVTEATITGLEPGETYITIVIALKNNQSEPLIGRKKTVQKTPFVTHPGYDTGNGIQLP 1030
QY 1087 AFGNSIPSKLSTLTLEAGYQOTN-----YLFPRTGEDKV-----WVAHH 1126
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1031 GTSQQSPVQCMIFESHGFRRTTPPTATPIHRPRPY-PNVGQALSQTTISWAPFQ 1089
QY 1127 GYTDY-----GT---AAQFMRPQOSNTQLTKITLIWDANYCVVVQ 1165
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1090 DTSEYIISCHPVGTDEPLQFRVGTSTSATLG---LTRGATYNIIVE 1135
```

RESULT 6

```
US-11-113-424-51
; Sequence 51, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; PRIOR FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 51
; LENGTH: 2715
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-113-424-51
```

```
Query Match 1.5%; Score 116; DB 7; Length 2715;
Best Local Similarity 18.9%; Pred. No. 4.8;
Matches 324; Conservative 176; Mismatches 511; Indels 704; Gaps 93;
QY 94 DTFGLGPEGLVAVDQPRDSTLOGINLGATP---TWTGYRSLRSHFSRLEYWQPTTG 150
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
990 DSIPIETQVL-----HEETIPGTDLKSLYSRAAGYKSVLKI-----TWTQA 1034
```

```
QY 151 KTDFLMIYSPDQGVHLL-----GKSPQARISNPSQTTQTQAW-----LLEASVS 194
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1035 VIPPNLM-----KVHLWAVVVGRLFKWFPAPSNLAYTFIMDKTDAYNKKVYGLSEAVVS 1089
QY 195 SRGEIYYQYRABDDTGCCEADBITHLQATQORYLHIVYGNRTASETLPGLDGSAPSQA 254
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1090 -----VGVEYES-----C-----LDLTLWEKRTA--VLQGYELDASNMG 1121
QY 255 DMLF--YLPFDY-----GERSNNLKTPPAPST-TGSMWLCRQDRFSEYEGFEIR 302
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1122 GMTLDKHHVLDVQNGILYKNGENGFISQOPPVVSSIMNG-----RRR 1165
QY 303 RLCRQVLMYHHLQALDKITEHNGPTLVSRLLIL--NYDESAIASTLVFVRRVGHQDGNV 360
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1166 SI-----SCPSQNGQADGNKLLAPVALACGIDGSLYVGDENYVRI--PFSGNV 1212
QY 361 VTLPPLELAYQDPSPPRHHAWQPMQVLANFNAIQRWQLVDLKGELPGLLYQDKGAWYR 420
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1213 TSV--LELENKOP--RHSS-----NPAHRYLAT--DPVTGDLVY----- 1246
QY 421 SAQLGEIGSDAVTWEKMOPLSV--IPSLQSNASLVLDINGDQDQDQDQDQDQDQDQ 478
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1247 -----SDTNRIYRPKSLTGAKDLTKNAEVVAGTGE----- 1278
QY 479 PDGSMTRFTPLNALPVEYTHPRAQLADLMGAGLSLVLIGPKSVRLYANTRDGFAGK-- 535
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1279 -----QCLPFD-----BARCGD--GGKAVEATLMSPKGMAIDKNGLIYFVDTMI 1321
QY 536 KDVVQSGDITLPVPGADPRKLVAFSDVLGSGQAHVSEVATKVTG-----WP- 582
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1322 RKVDQNGII-----STLGSNDL--TSARPLTCTDTSMHISQVRLWEPT 1362
QY 583 -----NLGRGRFCQPI--TLPGFSQA-----TEPNPAQ 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1363 DLAINPMDNSIYVLDNNVVLQITENRQVRIAGRPMHCQVPGVEYVPGKHAVOTTLESAT 1422
QY 610 -----VYLADLDSGPTDLIYVHT-----NRLDIFLKNKGNGFA 643
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1423 AIAVSYSGVLYITETDEKKINRIQVTTDGEISLVAGISECDCKNDANCDQYQSGDYA 1482
QY 644 EPTVLPFPGLRPDHTCQLOMADVQGLVASLILSVPHMSPHHWRCDLTNKMFWLLNEMN 703
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1483 KOAKLAPSSLAASPDGTLYIADLGNIRAVSKNKP-----LLNSMN 1525
QY 704 -----NNMGVH-----HTLAVRSS 717
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1526 FYEVASPTQOELYIPDINGTHQVTVSLVTGDIYNFYSYNDNDVTAVTDSNGNTLIRRD 1585
QY 718 SQ-----FWLDEKAAALTTGQTPVC-----YLPFFPIH-----TLWQT 749
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1586 PNRMPVRVSPDNQVILWLTIG-TNGCLKSMTAQGLEVLVFTYHNGSGLLATKSDEGTWT 1644
QY 750 ETEDEISGNKLVTTLYARG----- 769
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1645 PFYDSEGR--LTNVTFPTGVVTLNHLGDMDKAITVDIESSSREEDVSITSNLSIDSFYT 1702
QY 770 -----AWDCRERERFGVVE--QTDSHOLAQGNAPERTPPALTKNWYATGLP 815
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1703 MVQDQLRNSYQIGYDGLSLRIFASGLDSHYEPHVLG-GTA--NPTVAKRNTLPG-- 1756
QY 816 VIDNALSTEYW--RDDQAFAGFSPRFTTWQDNKDVPLTPEDDNRSYWFRNALK--GQLLR 871
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1757 --ENGQNLVWEWRPKQEA-----QKGVNV-----FGRKLAVNGRNL 1791
QY 872 SELYGLDDSTNKVPTVTVEFRSQVRLQHTDSRYVPLWSSVVSRYHYERIASPQCS 931
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1792 SVDFDRTTKTEK----IYDHRKFLRLIAYDTSGHPTLM-----LPSSKLMA 1834
QY 932 QNTLTSSDRFGQ-----PLKQLSVQYPRQOPAINLYPD-----TLDPK---LLANSY 976
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1835 VNTYIST--GQIYASTQRTTSEKVDYDSQGRIVSRVFDAGKTWSTYLEKSVLLHLS-- 1891
QY 977 DQQRQLRLTYQOSSWHHL-----TNTNTRVLGL-----PQSTRSDIFTYGA 1018
```

Db 1992 ---ORQIFEDM---WRLSALTWPSVARHTMQTIRSIGYRNINYPESNASIITYNE 1946
 Qy 1019 ENVPAGLNLLELLSDKNSLIADDPREYLGQOKTAYTDGQNTTLPQTPTQALIAFTETT 1078
 Db 1947 E-----GLLQ-----TAPLG-----TSRRVLPKYRQTRLSLSEILYDSTR 1981
 Qy 1079 V---FNOSTLSAFNGSIPSDKSLTLEQAGQOQNYLFPRT-----GEDKWVAHGTIDY 1131
 Db 1982 VSFTYDETAGVLKTVNLQSDGFICTIR---YRQIGPLIDRQIFRSEDGMVNAFDS-Y 2037
 Qy 1132 GTAAQFWRPQKSN-----TOLTGKI-----TLIMDANY-----CVVQTR-- 1167
 Db 2038 DNSFRVTSMQVINETPLPIDLYQFDDISGKVEQFGVGVYIYDINQIISTAVWYTKHP 2097
 Qy 1168 DAAGLTTSKYD-WRFILTPVOLTDINDNQHLITL--DALGRPITLRFMTGTEGNTGYSS 1224
 Db 2098 DAHGRKEIQEIPRSL-----MYWITIOYDNMGR-VTKR-----EIKIGPPAN 2140
 Qy 1225 PEKASFPSPDVAIAELKKPLPAQCOVYAPESWMPVLSQKTFNRLAEQWQKLYNARI 1284
 Db 2141 TTKYAYE--YVDVQGLQ-----TVYLNKIM-----WRYNYDL-- 2171
 Qy 1285 ITEDGRICTLAYRWQSQKAIPLISLNNPRLPHSLTLTTRDYHDHDEQEIQQVW 1344
 Db 2172 ---NGN-----LHLN-----PSSARKLTPIRYD----- 2192
 Qy 1345 PSDGFGRLQAAAR-HEAGMARQNE-----GSLIINVQHTENRMAVTRTEYDNKGQ 1397
 Db 2193 LRDRITRLGVDQVRLDSDGFLRQGTETFEYSSKGLLTVYKSGSGWTVIYR--YDGLGR 2250
 Qy 1398 PIRT-----YQPIFLNDWRY-----VSDASARQEKAYADT--HVY----- 1431
 Db 2251 RVSSKTSIGLQHPFYADLTPTTRITHVYNSHSSEITSLYYDLOGLHFAMEISSGDBFYI 2310
 Qy 1432 --DPIGREIKVITAKG-WFRETLTPMFTVNEDN 1463
 Db 2311 ASDNTGTPLAVFSSNGLMLKQOTAYGEIYFDSN 2345

RESULT 7
 US-10-858-730-72
 ; Sequence 72, Application US/10858730
 ; Publication No. US20050255568A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bailey, Richard B.
 ; APPLICANT: Blomquist, Paul
 ; APPLICANT: Dofen, Reed
 ; APPLICANT: Driggers, Edward M.
 ; APPLICANT: Madden, Kevin T.
 ; APPLICANT: O'Leary, Jessica
 ; APPLICANT: O'Toole, George
 ; APPLICANT: Trueheart, Joshua
 ; APPLICANT: Walbridge, Michael J.
 ; APPLICANT: Yorgey, Peter S.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
 ; FILE OF INVENTION: PRODUCTION
 ; FILE REFERENCE: 14184-030001
 ; CURRENT APPLICATION NUMBER: US/10/858,730
 ; PRIOR FILING DATE: 2004-06-01
 ; PRIOR APPLICATION NUMBER: US 60/475,000
 ; PRIOR FILING DATE: 2003-05-30
 ; PRIOR APPLICATION NUMBER: US 60/551,860
 ; PRIOR FILING DATE: 2004-03-10
 ; NUMBER OF SEQ ID NOS: 364
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 72
 ; LENGTH: 1192
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-858-730-72

Query Match

1.5%; Score 115.5; DB 6; Length 1192;

Best Local Similarity 21.5%; Pred. No. 1.5;
 Matches 167; Conservative 84; Mismatches 288; Indels 237; Gaps 42;
 Qy 293 YEXGFIRRLRCQVLMYHHLQALDSKITERHGPT-----LVSRLI-----LNYD 338
 Db 443 YEDGDGPESRFAKTMALVAEHAVALTIDEGQARTACKKVEIAERLINDITGNMGVD 502
 Qy 339 ESAI-ASTLVFVRVHEQ---DGNVVTLPLLELAYQDFSPRHHAHQWMDV-----LAN 389
 Db 503 ESSILDTLFTTATQGESESRDG---IETIE-AIRELKKRH-----PDVQTTGLSN 551
 Qy 390 FNAIQRWQLVDLKGEGPLGLYODKGAWYRSQRIGESDAVTWEKMOPLSVIPSLOS 449
 Db 552 ISP-----GLNPAARQVILNSVFLHCOEAG-LDSAIHVASKIILPMRIPREOR 598
 Qy 450 NASLVDINGDQLDWITVTPGLRGYHSQRDPGSGWTRFTPLNALPVEYTHPRA-----QL 503
 Db 599 NVA-----LDLV-----YDRRRED-----YDLOELMRLPEGVSAASSKEDRL 636
 Qy 504 ADLMAGLSDLVLIGPKSVRLYANTRDGPFAKDKDVQSGDITLTPVGADPRKLV----- 557
 Db 637 AELAGLPLFE-----RLAQRIVDGERNGLD-----ADLDEAMTQKPLQIINEHLLA 683
 Qy 558 ---AFSDVLGSGOAHL-----VEVSATKVTCPNLCGRGFGOPITLPGFSQAPATEPNA 608
 Db 684 GMTVGLFSGGQQLPFVLSQSAEVMKAAVA-----YLEPHMERSDDDSGKG 730
 Qy 609 QVYLADLDGSGPTDLIYVHTNRDLIFLNKSGN-----GFAEPVTLRPEGLRFDHTCOLQ 663
 Db 731 RIVLATVKG---DVHDIKGNLVDIILSNNGYEVVNIIGIKQIATIL-----EVAEDK 779
 Qy 664 MADVQGLGVLASLILSVPHSPHWRCDLTNNKFWLLNEMNNMGVHTLRYRSSSQFWLD 723
 Db 780 SADVVGMS-GLLVKSTVVMKEN-----LEEMNTR-GV-----AEKFPVL 816
 Qy 724 EKAALTTGQTPVCYLFPFPIHTLMQTE---TEDEISGNKLVTLRYA-RG-----A 770
 Db 817 LGGAALTRS-----YVENDLAEIYQGEVHVARDAFEGKLMDTMSAKRGEAPDENSPEA 871
 Qy 771 WDGRERFRGFGVVEQTDSSHQ-----LAQGNA-----PERT-----PPALTKNW 809
 Db 872 IKAREKE-----AERKARHQRSKRIAAQKAAEPEVEPERSDVAADIEVPAFPFGSR 925
 Qy 810 YATGLPVID-----NALSTEYWRDQAFAGSPRPTTQDNDKDVLTPEDDNSRWTFN 862
 Db 926 IVKGLAVADYTGLLDERALFLGWLGRGQGGGPSY-----EDLVETEGRPRLRYWLD 979
 Qy 863 RALKGQLL--RSELYGLDDSTN-----KHVPYTVTFERSQVRRLQHTDSRYPVLWS 911
 Db 980 RLSTDGILAHAAVYGYFPAVSEGNDIVLTPKPDAPVRYRFFHFPQOR--GRFLCI-A 1036
 Qy 912 SVVESRNYHYERIASDPQCSQNTLSSDRFGQPLKLSVQVPRRQOPAINLYPDTL 967
 Db 1037 DFIRSRELAARGEVDVLPQLVTM-----GQPIADFA-----NELFASNAYRDYL 1082

RESULT 8
 US-10-485-517-145
 ; Sequence 145, Application US/10485517
 ; Publication No. US20050256299A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Sheffield
 ; APPLICANT: Biosynex Incorporated
 ; APPLICANT: Foster, Simon
 ; APPLICANT: Mond, James
 ; TITLE OF INVENTION: Antigenic Polypeptides
 ; FILE REFERENCE: P100629MO
 ; CURRENT APPLICATION NUMBER: US/10/485,517
 ; CURRENT FILING DATE: 2004-02-02
 ; PRIOR APPLICATION NUMBER: GB 0118825.9
 ; PRIOR FILING DATE: 2001-08-02
 ; PRIOR APPLICATION NUMBER: GB 02000349.9
 ; PRIOR FILING DATE: 2002-01-09

Db 2986 WYKRGSLPARHQTHSQLRLHHVSPADSGEYVCRAGGPGPQEQEASFTVTVPSPSSSY 3045
QY 686 HWRCDLTMKPK-----WLL--NEMNNMGV----- 708
Db 3046 RLRSVPISIDPPSTVQGGQDASPKCLHDGAAPISLEWTRNQELEDNVHISPNIGSIIT 3105
QY 709 -----HHTLYRSSQFWLDEKAAALTTGQTPVCYLPPPIHTLW-----QTETEDEIS 756
Db 3106 IVGTRPSNHGTIRCASNAYGVAQSVNLSVHGPPPTSV-LPEGPVWVKVGAATLECVS 3164
QY 757 GNTKLVTTLRVARGAWDGRERFGYVEQTDH-----QLAQGN 796
Db 3165 AGEPRSSARWTRISSTAKLEORTYGLM-----DSTVVLQISSAKPSDAGTVCLQAQNALGT 3221
QY 797 APERTPPALTKNMYATGLPVI---DNALSTE---YMRDDQAPAGSPRFT-----T 841
Db 3222 AQKQVEIVDTGAMAPQVQABEALTEAGHTATLRCSATGSPARTIHWSKLRSLPLP 3281
QY 842 WQ-----DNKDVPLTPEDDNRYSYWFNRALKQGLRSELYGLDDSTNKHV---PY--TVTE 891
Db 3282 WQHRLEGDTLIIIPRAVQODSGQYICNAT-----SPAGHAEATIIILHVESPPYATTVPE 3334
QY 892 FR-----QVRLOHTDSKYPVLWSSVVE-----RN--YHVERIASDP---QCS 931
Db 3335 HASVQAGETVOLQCLAGTLPFTQMSRVSSLPGRATARNELLHFERAAPEDSGRYRCR 3394
QY 932 -QNITLSSDRFGQPLKQ-----LSVQYPRRQOPAINLYPDTLPDKLLANSYD----- 977
Db 3395 VTNKVGSAAFAQLLVQPPGSLPATSIAGSTTPTVQVTPQ-LETKSGISAVRPHCAVPS 3453
QY 978 DOORQLBLTQQSSW-----HHLTNTVTVLGLPDSRSTRDIFTY 1016
Db 3454 DRGTQLR-----WPKEGGLPFGHSGVDGLRQNLQDSQCG---TY 3492

RESULT 10

US-10-485-517-198
; Sequence 198, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynex Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; PRIOR FILING DATE: 2004-02-02
; PRIOR FILING DATE: 2001-08-02
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 198
; LENGTH: 805
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-198

Query Match 1.4%; Score 114; DB 6; Length 805;
Best Local Similarity 18.2%; Pred. No. 1;
Matches 158; Conservative 93; Mismatches 265; Indels 350; Gaps 42;
QY 842 WQDNKDVPLTPEDDNRYSYWFNRALK--GQLRSELYGLD-----DSTNKHV-----PYTV 889
Db 13 WQD-----YWDENKTFKNDNLGQKKFYALDMFPYPSPGAGLHVHGEYTA 58
QY 890 TEFSSQVRRLQHTDSKYPVLWSSVVEGRNHYRIASDPQCSQNTLSSDRFGQPLKQ-- 947
Db 59 TDIISRYKRWQGVNVLPMGH-----DAFGLPARQYA 90
QY 948 LSVQYPRRQOPAINLYPDTLPDKLLANSYDDQORQLRLT-----YQSSW----- 992

Db 91 LDTGNDPREFTKKNIOFTFKQIKELGFSY--DWDRVNTTDPYKYKWTQWFIQLYNKGLA 149
QY 993 --HHLTNTVTVLGLPDSR---SDIFTYGAENV---PAGGLNLELSDKNSLIAD--- 1040
Db 150 YVDEVAVNWCPCALGTVLNSNEEVIDGVSERGHPYRKPMPKQWLVKITETAYADQLLADLDDL 209
QY 1041 -----DKPREYLQOKTA-----YTDGQ-----NTTLPQTFTQALIAFTTETTVFNQ 1082
Db 210 DWPESLKDMQNRNIGRSEGAKVDFDNTGKVEVFTTRPDTIYGASFVLSPEHALVNS 269
QY 1083 STLSPF-----NGSIPSDKLSITLEQ-----AGYQOTVYLPPTGCD-KVWVAHH 1126
Db 270 ITTDEYKEKVKAYQTRASKKSLERTDLAKDGSVFTGAYATN---PLSEKGVQIWLADY 326
QY 1127 GYTDYGTAAQFWRPQKSNQTLGKITLIWDANCYVVVQTRDAAGLTTSKAYDWRFLTPV 1186
Db 327 VLSYGTGA-----IMAVPAHDDRDDEYFAKFDLPILIEVI 361
QY 1187 QLTDI-----NDNQHLIT--LD-----ALGRPITL----- 1209
Db 362 ECGNVBEAAVTGEGKHINSGLDLENEAAITKAIQLLEQKGAGEKKNYKLRDLWLFSRQ 421
QY 1210 RFNGT-----ENGKWTGYSSBEKASFSPSPD-----VNAALIEL-- 1242
Db 422 RYNGEPPIVTHWEDGTMTTVPBELLPLLLPETDEIKPSGTGESPLANIDSFVNVVDEKGT 481
QY 1243 -----KKPLPVAQCQVYA-----PES---WMPV-----LS 1264
Db 482 MKGRRETNTPQWAGSCWYLYRVIDPKENWNLADPEKLEHLPVDLYIGGVEHAVLHLY 541
QY 1265 QKTFNRL-----ABQDMQKLYNARIITEDGRICITLAYRRWVQSQKAIPLISLNNG 1316
Db 542 ARFWHKVLYDLAIVPTKEPQKLFNQGMILGEG-----NEKMSKSKGNVINPD 589
QY 1317 PRLPPHSLTLTDRYDHPDQEQIRQOVVPSDGGRLQLQAARHAGHAROR----- 1367
Db 590 DIVQSHG-----ADTLRLYEMF---WGPLDAAIAWSEKGLDGSRRFLDRVWRL 634
QY 1368 --NEDGSLIINVQHTENR-----WAVTGRTYDNKGQPIRTYQPYFLNDWRVVSNDSA 1418
Db 635 MVNEDGTLSSKIYTTNKSLLDKVYNQTVKKVTE-----DFETLGNNTA 677
QY 1419 RQB-----KEAYADTHYDP-----IGREIKVITAKGWFRTTLFTPMFTV 1458
Db 678 ISQLMWFINECYKVDEVYKPIEGFVKMLAPIAPHIGEEL--WSKLGHESSITVQPNPTY 735
QY 1459 NE-----DENDTAAEVK-----KVQM 1474
Db 736 DEALLVDDVEIVVQVNGKLRRAKIKI 761

RESULT 11

US-10-858-730-73
; Sequence 73, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000

APPLICANT: Zhang, Jie
APPLICANT: Zhou, Ping
APPLICANT: Cao, Yi-Cheng
APPLICANT: Drmanac, Radjoe T.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 785CIP4CN
CURRENT APPLICATION NUMBER: US/11/000,463
CURRENT FILING DATE: 2004-11-29
PRIORITY APPLICATION NUMBER: 10/291,265
PRIORITY FILING DATE: 2002-11-08
PRIORITY APPLICATION NUMBER: PCT/US01/02623
PRIORITY FILING DATE: 2001-01-25
PRIORITY APPLICATION NUMBER: 09/922,279
PRIORITY FILING DATE: 2001-08-03
PRIORITY APPLICATION NUMBER: 09/491,404
PRIORITY FILING DATE: 2000-01-25
PRIORITY APPLICATION NUMBER: 09/617,746
PRIORITY FILING DATE: 2000-07-17
PRIORITY APPLICATION NUMBER: 09/631,451
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: 09/633,870
PRIORITY FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 250
LENGTH: 1907
TYPE: PRT
ORGANISM: Homo sapiens
US-11-000-463-250

Query Match 1.4%; Score 113; DB 7; Length 1907;

Best Local Similarity 18.0%; Pred. No. 4.7;

Matches 296; Conservative 181; Mismatches 530; Indels 640; Gaps 80;

17 GGGAITGGEA---LPTGPD-----GMAALS-----LPLPISAGGYAPAFPTLVNS 61
589 GVGVTFTIETARTAQSPAPPQKVMCSMGSTTVRVSWVPPADRSNGVITQSVAYEA 648
62 GAGN-----SPFLGWDNCNMVTRRTHFGVPHYDETDFLGPGE-VLVWADQ- 109
649 VDGSDRGHRVVDGISREHSSWD--LVGLEKTEYRVVWRAHTDVGPGEPPVLRVTDED 706
110 ----PRESTLQGINLGATFT-----VTGYR-----SRLESHFSR----- 140
707 VPSGPRKVEVEPLNSTAVHYVWKLVPFSKOHGQIRGYQVTVYRLENGEPRGLPIIQVM 766
141 --LEYWPKTKTKDFTWLIYSPDQGVHLGKSPQARISNPSQTTQTAQWLLEASVSSRGE 198
767 LAEQWREE-----SEDYETISGLTPETYS-----VTVAAYTTKGD 805
199 QIYYQYRAEDDTGCEADEITHLQATQRYLHIVYGNRTASETLPG-----LD 247
806 GARSKPKIVTTTGAVPGRPMTMISTAMNTALLQWH-----PPKELPGLLGYRLQYCRAD 861
248 GSAPSQADWLPLYVDFDYGERSNNIKTPPAFTTG-----SWLCRQDRPSR-----YEVGFE 298
862 EAREN-----TIDFG-----KDDQHFTVTGLHKGTTVIFRLAAKNRAGLGEPEKE 907
299 IRT--RLRCROVLMYHHLQALDSKITEHNGPTLVSRLILNYDESALASTLVFVRVHGEO 357
908 IRTPELPSGPPQNLHVLTGTTSTTE-----LAWPPVLA-----ERN 945
358 GNVVTLPLELAYQDFSPRRHAHQMPDMVLANFNAIQRWQLVLDLKGELPGLLYQDKGAW 417
946 GRIIS---YTVVFRDINSQOE-----LQNIITDTRFTLTGLK-----PDTTYDIKVA 990
418 WYRSARQLGEIGSDAATWKKQPSVLPSSQSNASLVNDINGDGLDWITGFLRGYHSQ 477
991 W-----TSKSGSGLS--PSIQS----- 1005
478 RPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGKSVRLYANTRDGFAKGD 537
1006 -----RTMPEQVEPAK-----NFRVAAMKTS 1027

QY 538 VVSGDITLTPVGADPRKLVAFSDVLGSGQAHVLVEVSATKVTCWPNLGRGRFGQPIITLPG 597
DB 1028 VLLSWEVPSYKSAVPPK-----ILYNGQS--VEVDGHSNR-----KLIA 1065
QY 598 FSPATFENPAQVYVLADLDCSGPTDLYVHTNRDLIFLNKSGNGFAEPVTLR-----649
DB 1066 DLQNTETYS-----FVLMNR-----GSSAGLQHLHVSIRTPADLLPH 1102
QY 650 -----FPREGLRFPDHTCQLQMDVQGLGVASILSVPHMSPHHWRCDLTNMKPWL-----698
DB 1103 KLPASAYIEDGRFD-----LSMPHVQD-----PSLVRWFYIVV 1136
QY 699 --LNMNNMNGVHHTLYRSSQFWLDEKAAALTGTGTPVCYLPFPPIHTLMQTEDEIS 756
DB 1137 VPIDRVGSGM--LTPRWSTPELELLEAEIQG-----GEEQRR 1175
QY 757 GNKLAVTTLR--YARGAWD-----GREERFGRGVYVEQTDHSHQAQGNAPERTPPALTK 807
DB 1176 RRRQERLKPVAAQLDVLTPETTLGDKKNYRGF-----YNRP--LSPDLISY 1220
QY 808 NYVATGLPVIDNALSTEYMRDDQAFAGFSRFTTQONKDVPLTP-----EDNSRYM-----860
DB 1221 QCFVLA-----SLKEPMDQKRYAS-SP-----YSDEIVVQVTPAQOQSEPEMLWTGP 1267
QY 861 -----FNR-----ALKQOLLRS-----ELYGLDDST--N 882
DB 1268 VLAVALIILIVIAILLFKRKRTHSPSKDEQSIGLKDLSLAHSGSDPVEMRLNLTQTCMR 1327
QY 883 KHPVYVTEFRSQVRRLQHTDSRYPLVWSSVVERVHYVERIASDQCSQNTILSSDRFG 942
DB 1328 DHPPIPTDLADNIERLKAANDGL-----KFSQYESI--DP-----GOQFT 1366
QY 943 QPLKQLSVQVPRRQOPAINLYP-----DTLP--DKLIANSYDQQRQLRLTYOQSS 991
DB 1367 WENSLEVNPKRYANVIAYDHSRVILTSIDGVPSGDIYANVIDGYRQ-----1417
QY 992 WHHLTNNVTVLGLPDSRSDIFTYGAEVNPAGNLLELSDKNSLIADKPREYLGQOK 1051
DB 1418 ----NAYIATQGPLPETMGDFWRVWWEQRTATVMMTRLEEKSRVKCDQ-----1462
QY 1052 TAYTDGQNTTPLQPTTQALIAITETTVFNOSTLSAFNGSIPSPDKLSTLEQACYQQTNY 1111
DB 1463 -----YWPARG-----TETGLIQVT-----LLDTVELATYTVRTF 1493
QY 1112 LPFRTEGDK-----VWVAHHGYTDYGT-AAQFWRPQKOSNTQLTGKITLIWDANY 1160
DB 1494 ALHKSSESKEKRELQFOFMAW-PDHGYVEYPTPLAFLRRVKACNPLDAGPMVVCAGV 1552
QY 1161 ----CVVV-----QTRDAAGLTTSKADWREFLTVPVQLTDINDNQHLLITLDALGR 1205
DB 1553 GRTGCFVIDAMLERMKHEKTVDIYGHVTCWRSQRYNM--VQTED----QVVFIEHALLE 1606
QY 1206 PITLRFWGTENGKWTGYSSPEKASFSPSDVNAIELKPLPVAQCQVYAPESWMPVLSQ 1265
DB 1607 AATCGH--TEVPARNLYAHQKLGQVPPGSESVTAMELE-----1642
QY 1266 KTFNRLAEQDWQKLYNARIITEDGRICITLAYRRVWQSKAIPOLISILNNGPRLPPhSLT 1325
DB 1643 --FKLLAS-----SKAHTSRPISAN-----LPCNKFKNRLVINM-----PYELT 1679
QY 1326 LTTDRYDHPDPEQIQQOVVFSDFGRLQLAAARHEAGMARQORNBGSLIINVQHTENRWA 1385
DB 1680 RVCLQPIRGVEGSDYINASFLDGY-----ROOKAVIATQ-----GPL--AEATEDFW- 1724
QY 1386 VTGETEDYDKQPIRTTQPYFLNDWRVYVNSDSARQ-----EKEAYADTHVYDPIG-----1435
DB 1725 ---RMLWEHNSTII-----VMLTKREMREKCHQYWPABERSARYQYFVVDPMABYNMPQ 1776
QY 1436 ---REIKVITAKGWFRRTL-----FTPW 1455
DB 1777 YILREFVTARDQSGSRTIIRQFQFTDW 1803

```
RESULT 14
US-10-647-956A-6
; Sequence 6, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: Bowen, David
; APPLICANT: French-Constant, Richard
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/10/647,956A
; PRIOR FILING DATE: 2003-08-26
; CURRENT FILING DATE: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-647-956A-6

Query Match      1.4%; Score 112.5; DB 6; Length 915;
Best Local Similarity 18.8%; Pred. No. 1.6;
Matches 129; Conservative 85; Mismatches 222; Indels 249; Gaps 34;

QY 478 RPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSLVLGPKSVRLYA-----NTRD 530
Db 328 RNDABATRFWRNKVAPENSYTDSLYQLISATGREMANIGQQNQLPSALPSDNNYTT 387
QY 531 GFAGKXDVQSGDIT-----LPVFGADPRKLVAESDVL 563
Db 388 NYTRSYSDHSGNLTQIRHSSPATONNYTVAITLSNRSGVSLTTLTDPNQV----DTL 443
QY 564 GSQAHLVESATKVTCPNLGRGCPITLPGFSOPATEFPAPQVYLADLDGSGPTDL 623
Db 444 FDAGGHQTSLLPQGTLLW---TPRGELKQVNGPG-----NEWRYDSNG----- 485
QY 624 IYVHTNRDLTFLNKSNGFAPVTLRPPEGLRDPDHTCQLQMAQDVQGLGVASL----- 675
Db 486 ----MRQKVSQEPQTNTTQQRVIYLP-GLEL-RTTQSNATTTEELHVLITGEAGRAQV 539
QY 676 -ILSVPHMSPHHRCDLTNKKPMLNEMNNMGVHHTLRYR-----SSSQFWLDEKAAAL 729
Db 540 RVL-----HWE-----SGKP-----EDVNNQ-----LRYSDNLIGSQLELDNQCII 579
QY 730 TTGQTPVCYLPFPPIHTLMQFTEDEISGNKLVTTLRVARGAWDGREREFRG---FGV--- 783
Db 580 SEEE---YYPFGTALMAANSQTEAS---YKTIIRYS-----GKERDATGLYYGYTRY 626
QY 784 -----VEQTDSDHQAQGN-----APERTPPALTQKNYATGLPVID 818
Db 627 QPWAGRWLSADPAGTIDGLNLYRMVRNPNVSLQDENGLAPEKGYTKEVNF----- 678
QY 819 NALSTEWRRDQAF--AGFSRFTTQDNKDQVPLTPEDDNRSYWFRNALK----- 866
Db 679 -----DELKFKLAASKSHVVKWNEKE-----SSYTKNKSLLKVVVRVGDSDPS 719
QY 867 GQLLRSP--LYGLDDSTNKHVPYTVTFRSQVRRLQHTDSRYPVLWSSVRSRNYHYERI 924
Db 720 GYLLSHEELKGLIE-----KSGI-----IYSRL----- 742
QY 925 ASDPQCSQ-----NITLSSDRFG---QPLKQLSVQYPRROQPAINLYPD-----TLPD 969
Db 743 EENSLSLEKSKTNLSLGSISGYMARTIQDTISEYAEHKYRSN-HPDFYSETDFPALMD 801
QY 970 KLLANSYDDOORQLRLTYQSSWHLTNTNTRVLGLPDSRSDIPTFYGAENVAPAGLNLE 1029
Db 802 KSKBNDSYG-ERKIYAAWEKVYHDLKNK-----QSELHVNVALAHPTQLSNE 849
```

```
QY 1030 ----LLSDKNSLIADDKPREYLGQOK 1051
Db 850 ERALLQETEPATAIDREYNFKGVGK 874

RESULT 15
US-10-510-386-218
; Sequence 218, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 218
; LENGTH: 1432
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-218

Query Match      1.4%; Score 112.5; DB 6; Length 1432;
Best Local Similarity 20.1%; Pred. No. 3.3;
Matches 229; Conservative 147; Mismatches 408; Indels 357; Gaps 64;

QY 298 EITRRLCRQVLMYHHLQALDSKITEH-NGPTIVSLILNYDESATASTLVFVRRVHCQ 356
Db 73 QVDTPKKQAEKNAKSLSAKTEYQKRSVVLSALRVKADESQ-SDLKRYLKK--QEK 129
QY 357 DGNVVTLPPLLELAYQDFSPRHHAHQMPDMVLNFINAI-----QRQLVDLKGEGPLGL 410
Db 130 QGDV---KKIRSYIYVNGMAVHATKEVMEQVAAFPEVKVLPNKEKQLIKPTQSVKSA 186
QY 411 YQP-KGAWM-----VRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGQOLD 463
Db 187 AKDEKEIWNINRVDAPKAKLGYDGSQTVV-----ASI-----DTGVQ 225
QY 464 WVTITPGL---RGYHSQRPDGSWTRFTPLNAL-----PVEY-THPRAQLADLMGAG 510
Db 226 W--DHAPLAKKRYGYPQHPDPQPNHEFSWYDAVSGASBPDDLEHGHVTHVTGTVGSEPDG 283
QY 511 LSDL-VLIGPKSVRLYANTRDGFPAKGDVVQSGDITLPVPGAD--PRKLVAFSDVLG--- 564
Db 284 QNIGVAPGAKWIAVKAFFSDDG-GTDEDLIAAGEMILAPKDKGNPHPEMA-PDVVNSW 341
QY 565 SQQAHLVESATKVTQW-----PNLGRGR--PGQITLPGFSOPATEFPAPQVYLADLD 616
Db 342 SGGAGIDEFYRDIVKAWRAAGIPPEFSAGNVDPANP-----GGPASIAPANP--YPEAF 393
QY 617 GSGPTDLIYVHTNRDLTFLNKSNGFAPVTLRPPEGLRDPDHTCQLQMAQDVQGLGVASLI 676
Db 394 ATGATDI-----DNRLADF-----SLOGPS--PYDET-----KPEISAPGV-NIR 430
QY 677 LSVPHMSPHHRCDLTNKKPMLNEMNNMGVHHTLRYRSSSQFWLDEKAAALTTGQTPV 736
Db 431 SSVPG-SGYQDGDGTSM-----AGPHVAAAAALIKQADSSITVDETEKILMETATPL 482
QY 737 CYLPPPIHTLMQFTEDEISGNKLVTTLRVARGAWDGREREFRGFGVVEGTDSHQLAQ- 795
Db 483 TDSKF-----TESPNNGYGHGLVNVFDAVSATVDGLGR-----ABGQ 519
QY 796 ---NAPERTPALP-----TKNWTATGLP---VIDN-----ALSTEYMRDDQAFAGSPRF 839
Db 520 VGKEGDKSPPALNHOEITFIYSGSETPLKADVEDDVSIISVKLSYKTDADHETIAAQ 579
QY 840 TTWQDNKD-----VPLTPEDDNRSYWFRNALKQQLLRSELYGLDDSTNKHVPYTVTFERS 894
```

Db 580 TSGDYKGTAEAAVTSEGTKLSYKWIOTDFGKNKTESKYVEPIS-----PAVTGYKQ 634
Qy 895 QVRELQHTDSRYPVLWSSVVESSRYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPR 954
Db 635 DFB-----NSASGWLTSQVKN-----NER-GIFQSGPNAAAS-----GKNVPATNLGPGY 679
Qy 955 RQOPAINLY--PDTLPKLLANSVDDQORQLRLTYQOSSWHHLTNNTVRLGLPDSTRSD 1012
Db 680 ESSANMNLMPVPSVP-----KNQKLYLTYKY--WRDIEED----- 713
Qy 1013 IFTYG-----AENVPA-----GGLNLELLSDKNS----- 1036
Db 714 -FDYGFVVQPEGKEWIPAAEYSGKTSWKQGDLDLSEYGGQTIKWFNQLQSDSIEGD 772
Qy 1037 -LIADD-----KPREYLGQOKTAYT 1055
Db 773 GLYIDDDVALVKEVKSAGTKKRLGVEKQPAKMKDKTKKRMIDPKKAKPAEAL-QEK--T 828
Qy 1056 DGQNTTPIQPTTROALLIAFTET---TVFNQSTLSAFNGSIPSDKLSLTLEQAGYQOQNYL 1112
Db 829 ETKKAAPAVLFPVR-AQVSVLETKGSTYSNOAT-GAYSLAHAPGTYTLKAEAYGYESTAO 886
Qy 1113 FPRYGEDKVVVAHHGYTDYGTAAQFWRPQKQSNLTQLG-----KITLIWDA 1158
Db 887 V-KIESDKT-----TTADFVLKELKKGT-LTGTINKKKTGPVRHAKLYIIVEDA 933
Qy 1159 NYCVVQTRDAAGLTTSKAYDWMRELTVPQLTDINDNQHLITLDA-----LGRPI 1207
Db 934 -AVKPVQTDGGSYSLTA-YEGSYTVKVSANGYSSFSVDLKGDSKIDLDPPFIGYPG 991
Qy 1208 TL-----RFGWTENG---KWTGYSSPEKA-----SFPSPSDVNAAIE 1241
Db 992 EIGYDDGTGENAWAFYESGNGLAVKMTLENGQEKAMKGLGKFWDTFFDPGCGTDPFAVE 1051
Qy 1242 L 1242
Db 1052 V 1052

Search completed: December 16, 2005, 14:47:30

Job time : 31 secs

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:33:31 ; Search time 197 Seconds

(without alignment)
3126.297 Million cell updates/sec

Title: US-10-706-424-10

Perfect score: 7901

Sequence: 1 MNSQDSITELSLPKGGGA.....WFTVNDENDTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	4	US-10-706-424-10
2	7901	100.0	1474	4	US-10-754-115-45
3	6031.5	76.3	1476	3	US-09-817-514A-4
4	4581.5	58.0	1485	4	US-10-262-794A-32
5	4261	53.9	1506	4	US-10-754-115-49
6	4261	53.9	1506	6	US-11-020-848-2
7	3971	50.3	1428	4	US-10-754-115-60
8	3834	48.5	1493	4	US-10-753-901-18
9	3834	48.5	1493	4	US-10-754-115-18
10	3106.5	39.3	1444	4	US-10-609-113-11
11	3106.5	39.3	1444	4	US-10-754-115-40
12	1262	16.0	697	4	US-10-609-113-39
13	209	2.6	2386	4	US-10-156-761-7751
14	199	2.5	2060	4	US-10-381-596A-2
15	181.5	2.3	2364	4	US-10-156-761-7834
16	162.5	2.1	1250	4	US-10-156-761-7572
17	160.5	2.0	1385	4	US-10-282-122A-68242
18	160.5	2.0	1426	3	US-09-912-020-340
19	160.5	2.0	1426	4	US-10-282-122A-42617
20	160.5	2.0	1426	5	US-10-771-241-340
21	160.5	2.0	1551	4	US-10-437-963-181412
22	159.5	2.0	1329	5	US-10-450-763-56139
23	159.5	2.0	1329	5	US-10-450-763-58761
24	154	1.9	5215	3	US-09-861-289-2
25	154	1.9	5215	3	US-09-860-846-2
26	154	1.9	5215	3	US-09-988-384B-2
27	154	1.9	5215	3	US-09-836-821-2

ALIGNMENTS

RESULT 1

US-10-706-424-10
; Sequence 10, Application US/10706424
; Publication No. US20040103455A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens
; FILE REFERENCE: 62878
; CURRENT APPLICATION NUMBER: US/10706,424
; CURRENT FILING DATE: 2003-11-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1474
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-706-424-10

Query Match	100.0%;	Score 7901;	DB 4;	Length 1474;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1474;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MNSQDSITELSLPKGGGAITGMEALTPGPDGMAALSPLPLPISAGRGYAPAFITLNN	60	
Db	1	MNSQDSITELSLPKGGGAITGMEALTPGPDGMAALSPLPLPISAGRGYAPAFITLNN	60	
Qy	61	SGAGNSPFLGWDGNCVNTIRRRTHFGVPHYDETDITFLGPEGEVLVADQPRDESTLOGIN	120	
Db	61	SGAGNSPFLGWDGNCVNTIRRRTHFGVPHYDETDITFLGPEGEVLVADQPRDESTLOGIN	120	
Qy	121	LGATFTVTGYSRLESFHSFSLYKQPTTKTDFWLIYSPDGQVHLKSPQARISNPSQ	180	
Db	121	LGATFTVTGYSRLESFHSFSLYKQPTTKTDFWLIYSPDGQVHLKSPQARISNPSQ	180	
Qy	181	TTTQAOMLLASVSSRGEIYYQVRAEDDTGCCADEITHLQATQRYLHVYGNRTAS	240	
Db	181	TTTQAOMLLASVSSRGEIYYQVRAEDDTGCCADEITHLQATQRYLHVYGNRTAS	240	
Qy	241	ETLPLGDSAPSQADWLFYLVDFYDYSRNLKTPPAFTTGSWLCRODRFSRYEYGFPEIR	300	
Db	241	ETLPLGDSAPSQADWLFYLVDFYDYSRNLKTPPAFTTGSWLCRODRFSRYEYGFPEIR	300	
Qy	301	TRRLCROVLMYHHLQALDSKITEHNGTFLVSRLLNLYDESAIATSLVFRVRVGHQDGNV	360	
Db	301	TRRLCROVLMYHHLQALDSKITEHNGTFLVSRLLNLYDESAIATSLVFRVRVGHQDGNV	360	
Qy	361	VTLPPLSLAYODFSPRHHAHQMDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWMYR	420	
Db	361	VTLPPLSLAYODFSPRHHAHQMDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWMYR	420	

```
QY 421 SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGYHSORPD 480
DB 421 SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGYHSORPD 480
QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540
DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540
QY 541 SGDTITLVPQADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFCQPTITLPGFSQ 600
DB 541 SGDTITLVPQADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFCQPTITLPGFSQ 600
QY 601 PATEFPAPQVYLADLDGSGPTDLIYVHTNRDLIFLKNSGNGFAEPVTLRPEGLRDPDHTC 660
DB 601 PATEFPAPQVYLADLDGSGPTDLIYVHTNRDLIFLKNSGNGFAEPVTLRPEGLRDPDHTC 660
QY 661 QLMQADVQGLGVASLILSVPHMSPHWRCDLTNMKPMWLLNEMNNMGVHHTLYRSSSQF 720
DB 661 QLMQADVQGLGVASLILSVPHMSPHWRCDLTNMKPMWLLNEMNNMGVHHTLYRSSSQF 720
QY 721 WLDEKAAALTGTGTPVCYLPPPTHTLWQTEDEISGNKLVTLRYARGAWDGRERFRG 780
DB 721 WLDEKAAALTGTGTPVCYLPPPTHTLWQTEDEISGNKLVTLRYARGAWDGRERFRG 780
QY 781 FGVVEQTDHQAQGNAPERTPPALTKWYATGLPVIDNALSTERYWRDQAPAGFSPRFT 840
DB 781 FGVVEQTDHQAQGNAPERTPPALTKWYATGLPVIDNALSTERYWRDQAPAGFSPRFT 840
QY 841 TQDNKDVPILTPEDDNRSYWFRNALGQLLRSELYGLDDSTNKHVPYTVTFERSQVRRLLQ 900
DB 841 TQDNKDVPILTPEDDNRSYWFRNALGQLLRSELYGLDDSTNKHVPYTVTFERSQVRRLLQ 900
QY 901 HTDSRPVPLWSSVBSNHYERIASDPQCSQMITLSSDRFGQPLKQLSVQYPRROQPAI 960
DB 901 HTDSRPVPLWSSVBSNHYERIASDPQCSQMITLSSDRFGQPLKQLSVQYPRROQPAI 960
QY 961 NLXPDTLPDKLLANSYDDQORQLRLTYQSSWHLTNNTVRVLGLPDSRSDIPTTYGAEN 1020
DB 961 NLXPDTLPDKLLANSYDDQORQLRLTYQSSWHLTNNTVRVLGLPDSRSDIPTTYGAEN 1020
QY 1021 VPAGLNLLELLSKNSLIADDKPREVILGQOKTAYTDQNTTPIQTPTRQALIAFTETTVF 1080
DB 1021 VPAGLNLLELLSKNSLIADDKPREVILGQOKTAYTDQNTTPIQTPTRQALIAFTETTVF 1080
QY 1081 NQSTLSAFNGSIIPSDKLSITLBOAGYQOYNYLFPRTGEDKVVVAHGYTDYGTAAQFWRP 1140
DB 1081 NQSTLSAFNGSIIPSDKLSITLBOAGYQOYNYLFPRTGEDKVVVAHGYTDYGTAAQFWRP 1140
QY 1141 QKQSNQTLQTKITLIDWANCVVVQTRDAAGLTTSKADYDRFLTPVQLTDINDNOHLITL 1200
DB 1141 QKQSNQTLQTKITLIDWANCVVVQTRDAAGLTTSKADYDRFLTPVQLTDINDNOHLITL 1200
QY 1201 DALGRPITLRFWGTENGKMTGYSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM 1260
DB 1201 DALGRPITLRFWGTENGKMTGYSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM 1260
QY 1261 PVLISQKTFNLAEDQWOKLYNARIITEDGRICTILAYRWVQSOKAIPOLISLLNNGRPLP 1320
DB 1261 PVLISQKTFNLAEDQWOKLYNARIITEDGRICTILAYRWVQSOKAIPOLISLLNNGRPLP 1320
QY 1321 PHSITLTTDRYDHDPEQIRQQVVFSDGFRLLQAAARHAGMARQORNEGSLIINVQHT 1380
DB 1321 PHSITLTTDRYDHDPEQIRQQVVFSDGFRLLQAAARHAGMARQORNEGSLIINVQHT 1380
QY 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYYSNDSARQEKAYADTHVYDPIGREIKV 1440
DB 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYYSNDSARQEKAYADTHVYDPIGREIKV 1440
QY 1441 ITAKGWFRRITLTPFWFTVNEDENDTAAAEVKKVKM 1474
DB 1441 ITAKGWFRRITLTPFWFTVNEDENDTAAAEVKKVKM 1474
```

```
RESULT 2
US-10-754-115-45
; Sequence 45, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 1474
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-45
Query Match 100.0%; Score 7901; DB 4; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQNSQDSITELSLPKGGGAIITGMGEALTPTGPDGMAALSPLPIISAGRGYAPAFITLYN 60
DB 1 MQNSQDSITELSLPKGGGAIITGMGEALTPTGPDGMAALSPLPIISAGRGYAPAFITLYN 60
QY 61 SGAGNSPGLGWDNCNVTIRRTTHFOVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
DB 61 SGAGNSPGLGWDNCNVTIRRTTHFOVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
QY 121 LGATFTVTGVRSLRLESHFSLRYWQPKTKTDPMLIYSPDGOVHLGLKSPQARIINPSQ 180
DB 121 LGATFTVTGVRSLRLESHFSLRYWQPKTKTDPMLIYSPDGOVHLGLKSPQARIINPSQ 180
QY 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADBITHHLOATAQRYLHIVYGNRTAS 240
DB 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADBITHHLOATAQRYLHIVYGNRTAS 240
QY 241 ETLPGLDGSAPSQADWLFVLVFDYGRSNNLKTTPPAFTTGSWLCRQDRFSRYEYGFPIR 300
DB 241 ETLPGLDGSAPSQADWLFVLVFDYGRSNNLKTTPPAFTTGSWLCRQDRFSRYEYGFPIR 300
QY 301 TRRLCQVLMYHHLQALDLSKITEHNGFTLVSRLLINYDESAIASTLVFVRRVGHQDGNV 360
DB 301 TRRLCQVLMYHHLQALDLSKITEHNGFTLVSRLLINYDESAIASTLVFVRRVGHQDGNV 360
QY 361 VTLPPLLELAYQDPSPRHHAHQPMQVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWYR 420
DB 361 VTLPPLLELAYQDPSPRHHAHQPMQVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWYR 420
QY 421 SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGYHSORPD 480
DB 421 SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGYHSORPD 480
QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540
DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540
QY 541 SGDTITLVPQADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFCQPTITLPGFSQ 600
DB 541 SGDTITLVPQADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFCQPTITLPGFSQ 600
```


Db 1350 -----DTRMAIAERVDYDGTGAVIRSPQFFVLDNMYGAEV--SSSMATYI 1396
Qy 1430 VYDPIGREIKVITAKWFRTLTPFWTNEDENDT 1465
Db 1397 YDALARQLRMVNAKGYERTAFYFPWFNEDENDT 1432
RESULT 12
US-10-609-113-39
; Sequence 39, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; TITLE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 39
; LENGTH: 697
; TYPE: PRT
; ORGANISM: Paenibacillus apairius strain DB482
US-10-609-113-39

Query Match 16.0%; Score 1262; DB 4; Length 697;
Best Local Similarity 37.0%; Pred. No. 2.5e-90;
Matches 270; Conservative 119; Mismatches 291; Indels 50; Gaps 11;
Qy 736 VCYLPPPIHLMQTEDEISGNKLVITLYAYAGMDGRERFRGCVYEQTUSHQLAQ 795
Db 7 VCGVPPPIVVSIDHTVDEISGARIQKYTRNGVDYDRTDKFAGFGHIDTWEERDSEG 66
Qy 796 NAPERTPALTKNYATGLPVIDNALSTYWRDQAPAGFSP-RFTTWQ--DNKDVLPTP 852
Db 67 TLSISPPVLTRTYHTGQKQDERAVQVQWQDPAFQVPLVLTFFDQATQDVLDS 126
Qy 853 EDNSRYFWNRALKGQLRLRELY-----GLDDSTNKHVPYVTFEFSQVRLQHTDSRYP 907
Db 127 PNRREYWLRLSLRGMPLRNEIFAGDVVGLP-----PYQVESLRYQVRLMQSTDSECV 179
Qy 908 VLWSSVVESENHYERTASDPOCSQNTLSSDRFGQPLKQLSVQYPRQOPAINLYPDTL 967
Db 180 TL-PMQLEQTYNYEQASDPOCSQIQQWFDYGVAAQSIITIQPRRAQFEDNPYPHTL 238
Qy 968 PDKLLANSYDDQQRALATYQQSSWHHLTNVTVLGLPSTRSIDFTYGAENVVAGGLN 1027
Db 239 PDTSMSSYDSQQMLLTLTRQKAYHLADPEGWELNIPHTRLDSFIYADSVPAGIS 298
Qy 1028 LELSDKNSLIADKPREYLGQKQATYDQNTTFLQTPTRQALIAFTETTVNQSTLSA 1087
Db 299 AELLGGDTLRSALQAYGQGEIIVAGGE-----PDSRALVHYTRSAILDACLOA 352
Qy 1088 FNGSIPDKLSTLEQAGYQNTYLPRTGCEKVVAAHGYTDYGTAAQFWRQKQNTQ 1147
Db 353 YEGVLSDQNSLSSAGYQSRARILGSDGEADIFVAEQGFTRYADQNFRILGQOSSL 412
Qy 1148 LTGKITLWDANYCVVQTRDAAGLTTSKYDMRFLTPVQLTDINDNQHILITLDALGRPI 1207
Db 413 LTGEQVLTWDDNFCVTSIEDALGNQIQIAYDFRVEAIQITDANNVNVQSLDALGRVV 472
Qy 1208 TLRFWGTENGMTGYSPKASPPSDVNAALTELKPLVAQCQVYAPESWMPVLSQKT 1267
Db 473 YSRWTGTEGIEGTF-RPE-AEFSPPETMEQALALASPLPVASCCVYDAHSWMTITLQ 530

Qy 1268 FNRL---ABQDMQKLYNARIITEDGRICLTAYRRVQSQKAIPQLISLNNGPRLPHSL 1324
Db 531 LSAIVDPSEKQMSFLIANELIMPDGRIKIRARGAPMWLQRLLPAPAKLSEADRKPPHYV 590
Qy 1325 TLTTRYDHDPEQIIRQVVFSDGFRLLQAAARHEAGMARQNEGSLIINVQHTENRW 1384
Db 591 VLAADRYPDPSQIQASVVFSDGFRGTIQTAKR-----ADTRW 629
Qy 1385 AVTGRTEYDNKQOPRTYQPYFLNWRVYVNSDQKAEKAYADTHVDYDGRKIKVITAK 1444
Db 630 AITERIDYDETGAIVRSFQFFYIDDMWYVYGEAV--SGSMYATIIYYIDALARQLRMVNAK 687
Qy 1445 GWFRTLTPT 1454
Db 688 GYERTAFYTP 697
RESULT 13
US-10-156-761-7751
; Sequence 7751, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7751
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7751

Query Match 2.6%; Score 202; DB 4; Length 2386;
Best Local Similarity 18.3%; Pred. No. 1.7e-05;
Matches 344; Conservative 184; Mismatches 647; Indels 706; Gaps 89;
Qy 13 SLPKGGGATG--MGEALPTG---PDGMAALSLPLISAGR---GYAPAFTLNYS-- 61
Db 318 TLASGSSDAGDYRATSLTSTGWEVSTGSGAFTYVNPQLPKPMGSAFSLSLNSQS 377
Qy 62 -----GAGN---SPFGLGWDNCVMTIRRR-----THFGVPH-----YDETDTFLGPEGEVL 104
Db 378 VDGRTSASNNQASWAGMDLVGVYIERRYVNCSEDLPTIGMCDWSPNSAKEPSGAVY 437
Qy 105 VVADQPRDESTLQINLGATFTVTG---YRSRLSHFSRLYWPQTKTKT-DFWLIYSP 160
Db 438 VI-----NLNGVTSILQDNTGSGAYHLKNDPGWRVQLPDGCGAGSDGYWVISTQ 489
Qy 161 DGOVHLGLKSPQAR-----ISN-----PSQTTQTAOWLLLEASVSSRG 197
Db 490 DGRYFGWGRSERTATASVETEPVGNDDTGPCHDQFPPECTQAWRSLDRAVDANE 549
Qy 198 EQIYYQYRAEDDTGCEADEITTHLQATAQR-----YLHIVYTG----- 235
Db 550 VETMYFY-----DKEYNHVRSVANSKDAREVSSGVYKVIQYGWSSQIPDGKLP 599
Qy 236 -----NRTASE-----TLPLGLDGSAPQADWLFYLVFDYDGRSNNL--KT- 273
Db 600 KVLSHVNRCIERVBQBNPLRDEPATCPTFDKXTPSYDPVVDLMCDGTGSADYNCAGKY 659

QY 254 ADWLFYLVDFYGBRSNNLTPPAFTTSGWLCRODRFSRYEYFEIRTRRLCRQVLMYHH 313
Db 722 A-----LTGNELO--PAY-----SYIHTKDAFK-----HT 744
QY 314 LQALDSKITEHNGPTLVSRLL--ILNYDESA-----IASTLVFVRVG-----HQDGNVVT 362
Db 745 LTDVDVRNTN-----SGLGNVLNNEKANVIVKADTKDYMKYLSTDFPLDNSSDISVDK 797
QY 363 LPPLLEYAQDFSP-----RHAIHQW-----PMDVLANFNALQWQ 397
Db 798 LDHINKHYLSLSPADDTAIGNVRAMTNNELIRTSTVMNPSLYETTPSNVLKTRPVLNPIK 857
QY 398 LV-DLKGEGLPGLLYO-DKGAMVRS-----AORLGEIGSD-----431
Db 858 VIKXKXNSGRYLYIYEAPEGYKNQVKSITERSITQNTPEITFDIYNRGLTPTGTVS 917
QY 432 ---AVTWKMQPLSVSPISQSNASLVINDINGDQGLDWITOPGLRGYHSQRPDGSMTRFT 488
Db 918 IRYATTIWDENSEI-VRPTEEQSL-----HNNLELSYVIT-EDLSGN-----KKFVS 962
QY 489 LNALPVEYTHPRAQLADL-MGAGLSDL-----VLIGPKSVLYANTRDGPAGKDVQ 540
Db 963 VIDVPFKIALAKEYASTLTIGKDAASFPKQSDVDNGLG-ESVNLQNTAN-FTNSEGIK 1020
QY 541 SGDTILP-----VPGADPRKLVAFSD-----VLGSGQAHLEVESATKVTCPN 583
Db 1021 EIIVTIPKONIKTNLTALIPDTEKIRVYVTTDDVRNGVNSNPTDLTKRTAVYV-----1076
QY 584 LGRGRFGQP-----ITLPGFSQATEFNPAQVYLADLDG-----SGFTDLIYV 626
Db 1077 -----FDEPLVLNNGOSQPTNMVTPEDAPILTKAH-SQIFTKGLDNTWLSGNKVELET 1130
QY 627 HYNRLDI---FLKNSGNFABPVTLPPEGLRFDHTCQLQMAQVQGLGVASLLISVPHMS 683
Db 1131 EDNRGLVVKVYTNESGNTIQNSLTSGKKKNTXN-----VSPQOM-1170
QY 684 PHWRCDLTNMKPWLLNEMNMGVHHTLRYRSSQPMWLDKAAALTGOTPCVCLPF--741
Db 1171 -----IDRLN-----RHYKFRVDNQ--LDPTGHYAKGQTKVNLVIYE 1208
QY 742 -----PIHLMQT-----ETDEISGNK-----759
Db 1209 VPEGSVIADYKTTDGEVLSPLVTVNSQIBGTBYATPATIPDRVTFTETDDKVKVKTIS 1268
QY 760 --LVTLRYARGADWGRERFPGYVQDTSQHLAQGNAPERTPPAL-----805
Db 1269 YHLISTPENOSGTGVGKQITIEVHYV--EPITTYEQIPNDAPQETPVALETRYVDSEGNE 1327
QY 806 -----TKQWYATGLPVIDNALSTE-YWRDDQAFAGFSPRET-----840
Db 1328 VQTEBEGTHDAPGIIADKQYTOQTAAENGITTHVYQRIQSEIPNEAPQETPVALETCY 1387
QY 841 -----TWQDNKQVP-----LTPEDDNSRYWFNRLKGLLRSELYGLD 878
Db 1388 VDSEGNVEQTEBEGTHDAPGIIQDKWQYTOQTTEGITHYIQR-----IQSE---IP 1438
QY 879 DSTNKHVYVTVTFR-----SOVRRLQHTDSRYPLVMSVBSRNYHYERIASDQCQONIT 935
Db 1439 NEAPQETPVALETRYVDSEGNVEQET-----BEGTHQPPSIIGDKWQYTGOT 1486
QY 936 LSDFRFGQPLKQLSVQYPRQQAIPALNLYPDLTPDKLLANSY-DDQORQLRLTYQSSWHH 994
Db 1487 TTADGI-----TYVVERIQSEIPNEAPKETPIQLEVTYVDQEGNEVQET---BEGTHH 1538
QY 995 LTNNTVRVLGPDSTRSDIFTYGAENVPAAGLNLLELLSDKNSLIADDPREYLQ-QKTA 1053
Db 1539 APG-----IIG-----DKWQYTOQTTEGITHVYVERIQSEIPNEAPQETPQVBVTR 1587
QY 1054 YTDQONTPTLOTP-----TRQALIAFTETTVNQSTLSAFNCSIPSDKL 1097
Db 1588 YVNSEGNVEQTEBEGTHQPPSIIGDKWQYTOQTTEGITHVYVERIQSEIPNEAPKE-T 1646
QY 1098 STTLEQAGYQOTNYLFPRTGEDKVVVAHGYTDYGTAAQFWRPQKQSNQNTQLTGKITLWD 1157

Db 1647 PVQLEVTRYVDFTD-----GNEVQETEBGTHQPPGLIGDKM-----QYTGRTV-----1688
QY 1158 ANYCVVVQTRDAAGLTTSKYDMRFL-----TPVQL-----TTINDNQHLITLD 1201
Db 1689 -----EKDGITT---YVYERIQSAIPNEAPQETPVQLEVTRYVDITGNEVQETEE 1735
QY 1202 ALGRPITL-----RFGW---TENGMQTVSSPEKASPPSDVNAALIEKLKPLPVAQCQV 1253
Db 1736 GTHQRYIIGDKWRVSGVTVTENG-ITRHVY-ERIQSKVPNDA---POETPVQLEVTRY 1789
QY 1254 YAPESMMPVLQKTFNRLAE-----NEIQETTEGKHQPPGIIIGDRWQ---YTKGVTEKDG-IITYVYER-I 1835
Db 1790 VDPEG-----NEIQETTEGKHQPPGIIIGDRWQ---YTKGVTEKDG-IITYVYER-I 1835
QY 1301 QSKAIPQLISLILNNGPRLPHSLTLTTDRYDHDPEQIIRQQVVFSDGGRLLQAAARHE 1360
Db 1836 QSE--IP-----NNPQETPVELEVTR-----YVDGEGNEVQET-----1867
QY 1361 AGMARQBNEDGSLIINVOHTENRMAVTGR-TEVDNKGQPIRTYQPYFLNDRVVSNDAR 1419
Db 1868 ---TEGKHQPPSI---GDRMQVTGKVTESKG---ITTY--VYERIQSKVPNDAPR 1912
QY 1420 QSEKAYADTHVYDPIGRIKIVITAKGWFRTLTFTWFTVNEDENDTAAEVKK 1471
Db 1913 VDIDELKITIYVDYTNGREI-VPSRKQQLPPEQF-----IGQDWQYTGKHIEK 1958

RESULT 15
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 2.3%; Score 181.5; DB 4; Length 2364;
Best Local Similarity 16.8%; Pred. No. 0.00072;
Matches 274; Conservative 180; Mismatches 460; Indels 715; Gaps 77;

QY 7 FSITSLPKPGGGAITGMGEALTPTGP-----DGMALSLPLPISA-GRGYAPAPFTLY 59
Db 305 YATVSSSSDAGDY---RASTLSPTGSEVSTGSGSFNYSVPFIQVAPPWGSAPSLAMS 361
QY 60 NS-----GAGN---SPFGLGWDGNVMTIRR---THFGVPH-----YDETDFTLGP 100
Db 362 DSQVSDGRTSASNQASWSGMDLVNVTERRYNCTDGLPTIGDMCWDSPNSAKEPD 421
QY 101 GEVLVADQPRBESTLQGINLGTFTVTGVSRLSHFSLEVMQPKT-----TGKT-D 153
Db 422 GAAYVSLNGTTSSELIDNNGSGAY-----HLKNDPGRVQRULPDGHGANGGE 470
QY 154 FMLIYSPDQGVHLLGKSPQARISN-----PSOTTQTAOWLLE 190

Db 471 YWISTQDQRYFPGMGRSERTSTATASVPTVPVGVNDAGEPCHDQPEPCTQAWRNLD 530
Qy 191 ASVSSRGEQIYYQIRABDDTCCEADEITHLQATAQRYLHIVVYGNRTASTLPLGDSGA 250
Db 531 RAVDANEVETWIFY-----DKENHYRSVA----- 555
Qy 251 PSQADWLFYLVFDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEYGFETRRLCRQVLM 310
Db 556 -----NTDKAREYSSGVYKE-----IQYGM----- 576
Qy 311 YHHLQALDSKITEHNGPTLVSRLILNYDESAIAGTLVFRVRVGHQDQGNVTLPLLELAY 370
Db 577 -----SSQITDGKVPKVELSHV-----RCIERVQENDPLADEPAACPA 616
Qy 371 QDPSPRIHAWQPMVLANFNAIORMQVLKLGELPLGLYQDKGAWYSAQBLGIGS 430
Db 617 FDRKPGSYPD-VFVLDLMDGDSAD-----YNCAGKTYF----- 648
Qy 431 DAVTWERQWPLSVTPSLQSNASLVYDINGDGLDWITGPGLRGYSQRPDGSWTRFTPLN 490
Db 649 -----PTPFSTDMLDI----- 660
Qy 491 ALPVEYTHPRAQLADLMGAGLSLVLIQPKSVRLYANTRDGFAGKDVQSGDITLPPVG 550
Db 661 -----KTYVSDQG-----TGWDLVQYQNKYGMFN 686
Qy 551 ADPRKLVAFSDVLGSGQHLVEVSATKVTCPNIGRGRPGQITLPGFSOPATFNPQAV 610
Db 687 PD-----GTIGKTLW-----DYIQKRT 704
Qy 611 YLADLDGSGPTDLI--YVHTNRLDIPLKNSGPAEPVTLRFPPEGLRPHDHTCLOQADVQ 668
Db 705 Y-----GDG-DDIVLPVINFRDLD-NKVG-----AELNFR----- 736
Qy 669 GLGVASLILSVPHMSPHHRCDLTNMKPWLINEMNNMGVHHTLRYRSSQFVLDEKAAA 728
Db 737 -----IKEHGLGATTIVSYGFANACDIDHLPQ 766
Qy 729 LTTGQTVCYLPPFIHTLMQETEBEISGNKLVTTLRYARGAWGRERERFGYVEQTD 788
Db 767 ASNTQD-CY-----WQKWTE-----GETDSKTGWFKKF-LVTQV- 799
Qy 789 SHQLAQGNAPERTPPALTKNWYATGLPVIDNALSTE-----YWRDDQAPAGFSRFTWQ 843
Db 800 -----QVDPVTTN--QDGAPVMTTSYTEDGAGWHFTNDPLIKDEDESMTDW 846
Qy 844 DNKDVPILT-----PEDDSRYWFNRALKQLLSRLYGLDDSTNKHVPVTVTEPRSQVRL 899
Db 847 GYQEVQVTTGAGAQTKKSWLYRGLSG-----DRTSK----- 879
Qy 900 QHTDSRYPLWSSVVEGRNHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPPA 959
Db 880 -----ADASATKTVTV-DDGDG----- 895
Qy 960 INLYPDT--LPDKLLANSY-DDQORQLRLTYQSSWHHTNTVTUVLGLPDSRSDIFTY 1016
Db 896 -NNYTDSDLSGRILSTSLRDDTGTSHERTYHK-YWDH-----NTAQYDGLPDAR----FVR 946
Qy 1017 GAENVPAGLNLLELLSKNSLIADDKPVELYGOQKATAYDQONTT---PLQPTT-ROALI 1072
Db 947 EKE-----TTNTKVSSGW-REHTVETEYDDTEGASTTFLPMKTDWQSSV 993
Qy 1073 AFTETTVF-----NOSTLSAFNGSIPSKLS-----TTLEQAGYQOQNY----- 1111
Db 994 SDNRCYTGAYNTDNYDSTGAQRQVTLQDQVHYSGVSCSSIASDKQDGYASTLYDNATS 1053
Qy 1112 -----LPRTGEDK-VWVAHHGYTDYGTAAQFWRPQKQSNQTLTG-KITL 1154
Db 1054 VDANKPVDGNPTESTRYTKSGSYRSTW---SGYDDAGRV--MWSEGDGSHNRLLTKYSPAN 1108
Qy 1155 IWDANYCVWVOTRAAG-----LTTSAKYDWRFL-TPVOLTDINDOHLITLDALGRPI 1207
Db 1109 TWPLN-GVIVTTPDPDGPALPAHTALTSTAMTSRFGKPTS IQDANGNVTKMSLDAAGRLV 1167

Qy 1208 TLRFW-GTENGKMTGYSSPE-KASFPSPSPDVNAAIELKPLPVAQCQVYAPESWMPVLSQ 1265
Db 1168 EV-WRPTETG-----SSPSMKFSYTIPTSTNSA----- 1194
Qy 1266 KTFNRLAEQDWQKLYNARIITDGRICTLAYRRWVQSKAIPQLISLNNNGPRLPPHSL- 1324
Db 1195 -----GVPDV-----DGYPHVATHVLQ 1212
Qy 1325 TLTTDRYDHDPEQOIROQVVFSDGFRLLQAAARHEAGM-ARQNRNEDGSLIINYQHTENR 1383
Db 1213 SGITYLSSH-----AYVDGLGRARETQTPMGNGVDAAATGNEVPNRQVSVTRYDSA 1262
Qy 1384 WATGRTE-YDNKG-----QPIRTYQPYFLNDM--RYVSN----- 1415
Db 1263 GNVGTGSAVFRNOCGTAGSGGFSKAKVEDLPSYTDLVL-DWAGRAITSRLQVNGASQDAGR 1321
Qy 1416 -----DSARQEKAYADTHVYDPIGREIKVITAKGWFRRTLTPTWPTVNEDE---- 1462
Db 1322 VDTTYDGDFTSVKNKVDAAADTYT-DVYGQVSKVVEHTG--SATYTTAYTAKDELIKI 1378
Qy 1463 -----NDTA 1466
Db 1379 TDPGRNDTS 1387

Search completed: December 16, 2005, 14:47:04
Job time : 211 secs

This Page Blank (uspio)